

Db 123 GHOKEVGGDFLEAHMFSPALKAGASCKVTDENNGTYLVSTFLWEGQVSLVLLIHPSEG 182

Qy 186 VSALSNANQGYDRVFTGQFYNGTSQVHSEGLILNTNAELCOYLDNRDQEGYCVRPQ 247

Db 183 ASALARANQGYDRIIFKQGYNGTSQVHSEGLILNTNAELCOYLDNRDQEGYCVRPQ 242

Qy 248 HMPCAALFHTSKNKKVSVLSKOEKSPERSNGVGEIMKENTISVSKCN----- 297

Db 243 HMPCEALFHTSVSRNDISVLSKKNLPHRSKGVGEIM-KNQHIDVSOCKNKEVKECKQ 301

Qy 298 ----- 297

Db 302 IGMKIPVPGVYTGLOGRWLTFNCFNQIQLDTAKISOCKLKGKLIYLMGDSLTRQWIIYLPKVM 361

Qy 298 -TLKSVDLHESGKLOLAQVDDRNINIQMKYCYPLIGSMYSVKEMEYLTIRADITGG 356

362 KTLKFEDELHETGFKKKLLDLDAEHTQLOMKKSHFVYIQLFSVDHGVIPQEDIRLIG 421

Qy 357 ENKTVIVISLQCHFRFPFDIVFIRALNVHKAIOHLLRSPDTWVIKTNIREMYNDAE 416

Db 422 DKDVTIVITFQCHFRFPFDIVFIRALNVHKAIOHLLRSPDTWVIKTNIREMYNDAE 481

Qy 417 RESDFHGYIQYLIKDFKDLNAGVVDWMTIAYGTNNVHPQHVGVGNINILLNYIC 475

Db 482 REGDFHGYIQYLIKDFKDLNAGVVDWMTIAYGTNNVHPQHVGVGNINILLNYIC 540

RESULT 2

APB_HUMAN STANDARD; PRT: 4563 AA.

ID APB_HUMAN

AC P04114; O00502; Q13787;

DT 01-NOV-1986 (Rel. 03, Created)

DT 01-NOV-1986 (Rel. 03, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE APOLIPOPROTEIN B-100 PRECURSOR (APO B-100) [CONTAINS: APOLIPOPROTEIN B-48 (APO B-48)].

DE APOB.

GN Homo sapiens (Human).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RX MEDLINE-870161385; PubMed-3763409;

RX Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J., Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.; "Complete cDNA and derived protein sequence of human apolipoprotein B-100.";

RL Nucleic Acids Res. 14:7501-7503(1986).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE-870161385; PubMed-3652907;

RX Ludwig E.H., Blackhart B.D., Pierotti V.R., Calati L., Fortier C., Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.; "DNA sequence of the human apolipoprotein B gene.";

RN DNA 6:363-372(1987).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE-87008488; PubMed-3759943;

RA Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H., Gotto A.M. Jr., Chan L.;

RA "The complete cDNA and amino acid sequence of human apolipoprotein B-100.";

RT J. Biol. Chem. 261:12918-12921(1986).

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE-87041416; PubMed-3464946;

RA Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J., Lee N., Brewer H.B. Jr.;

RA "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino acid sequence.";

RT Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE-87161758; PubMed-3030729;

RA Cladaras C., Hadzopoulou-Cladaras M., Nolte R.T., Atkinson D., Zannis V.I.;

RA "The complete sequence and structural analysis of human apolipoprotein B-100: relationship between apoB-100 and apoB-48 forms.";

RL EMBO J. 5:3495-3507(1986).

RN [6]

RP SEQUENCE OF 709-906 FROM N.A.

RX MEDLINE-85270450; PubMed-3860836;

RA Deeb S.S., Motulsky A.G., Albers J.J.;

RA "A partial cDNA clone for human apolipoprotein B.";

RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).

RN [7]

RP SEQUENCE OF 3056-3159 FROM N.A.

RX MEDLINE-86041888; PubMed-3903660;

RA Mehrlabian M., Schumaker V.N., Fareed G.C., West K., Johnson D.F., Kirchgesner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;

RA "Human apolipoprotein B: identification of cDNA clones and characterization of mRNA.";

RL Nucleic Acids Res. 13:6937-6953(1985).

RN [8]

RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.

RX MEDLINE-86093680; PubMed-3841204;

RA Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O., Bjursell G.;

RA "Molecular cloning of human apolipoprotein B cDNA.";

RL Nucleic Acids Res. 13:8813-8826(1985).

RN [9]

RP SEQUENCE OF 3109-4563 FROM N.A.

RX MEDLINE-85300528; PubMed-2994225;

RA Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F., Uredea M.S., Levy-Wilson B., Powell L.M., Pease R.J., Eddy R., Nakai H., Byers M., Priestley L.M., Robertson E., Rall L.B., Besholtz C., Shows T.B., Mahley R.W., Scott J.;

RA "Human apolipoprotein B: structure of carboxyl-terminal domains, sites of gene expression, and chromosomal localization.";

RL Science 230:37-43(1985).

RN [10]

RP SEQUENCE OF 1-291 FROM N.A.

RX MEDLINE-86149325; PubMed-3513177;

RA Protter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V., Chen G.C., Kirsher S.W., McEnroe G., Kane J.P.;

RA "Isolation of a cDNA clone encoding the amino-terminal region of human apolipoprotein B.";

RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).

RN [11]

RP SEQUENCE OF 1-1670 FROM N.A.

RX MEDLINE-86287319; PubMed-3461454;

RA Yamanaka M., Hort Y.J., Hierrild K.A., Chen G.C., Kane J.P.;

RA "Analysis of cDNA clones encoding the entire B-26 region of human apolipoprotein B.";

RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).

RN [12]

RP PARTIAL SEQUENCE, AND IDENTIFICATION (APO-B48).

RX MEDLINE-88018019; PubMed-3659919;

RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H., Silberman S.R., Cal S.-J., Deslypere J.P., Rosseneu M., Gotto A.M. Jr., Li W.-H., Chan L.;

RA "Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon.";

RL Science 238:363-366(1987).

RN [13]

RP DOMAINS.

RX MEDLINE-87039351; PubMed-3773997;

RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr., Innerarity T.L., Blackhart B., Taylor W.K., Marcel Y., Milne R., Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;

RA "Complete protein sequence and identification of structural domains of human apolipoprotein B.";

RT

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	288	16.6	358	9	A1833131	A1833131 at75d06.x
C 2	169	10.5	905	12	BG199761	BG199761 RST19055
C 3	148	9.2	343	12	BG201797	BG201797 RST21139
C 4	112	6.9	112	9	A1581835	A1581835 ar34g08.x
C 5	111	6.9	946	12	BG217603	BG217603 RST37315
C 6	109	6.7	530	17	AC0439960	AC0439960 HS 5059.B

SUMMARIES						
Result No.	Score	Query		Length	DB ID	Description
		Match	%			
C 1	268	16.6	358	9	A183131	A183131 at75d06.x
C 2	169	10.5	905	12	BG199761	BG199761 RST19055
C 3	148	9.2	343	12	BG201797	BG201797 RST21139
C 4	112	6.9	112	9	A1581835	A1581835 ar34g08.x
C 5	111	6.9	946	12	BG217603	BG217603 RST37315
C 6	109	6.7	530	17	A0439960	A0439960 HS 5059.B

seq primer:

IMAGE Consortium (info@imageconsortium.org)

IMAGE Consortium (info@imageconsortium.org)

Fri Oct 12 16:09:17 2001

us-09-729-454-1.rsp

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

Protein search, using sw model

Run date: October 12, 2001, 16:02:37 ; Search time 16.9 Seconds
(without alignments)

962,802 Million cell updates/sec

Title: US-09-729-454-1

Sequence: 1 MKISMINKSLALLFLILAS.....VHPPOHVVGNQINILNYIC 475

Scoring table:

BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SUMMARIES

Result No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Match	Length	DB ID	Description
1	1606	64.6	540	1	BRL1_RABIT
2	107.5	4.3	4563	1	ABP_HUMAN
3	104	4.2	694	1	VGIL_HSV6U
4	103	4.1	793	1	STL_MYCPN
5	101.5	4.1	1902	1	P2P_LACPA
6	101.5	4.1	3649	1	ACVS_NOCIA
7	100	4.0	1031	1	TERT_EUPAR
8	98.5	4.0	802	1	CSD2_ECOLI
9	98	3.9	1174	1	NOS_RHOPR
10	97.5	3.9	271	1	NPH1_RAT
11	97.5	3.9	623	1	ABF2_HUMAN
12	97	3.9	988	1	P1NH_HUMAN
13	96.5	3.9	773	1	GLGB_SYNP7
14	96	3.9	770	1	RPOD_ASHMT
15	95.5	3.8	1102	1	RSD1_ECOLI
16	95	3.8	819	1	POIX_TORAC
17	95	3.8	1328	1	GUAD_MOUSE
18	93.5	3.8	454	1	DNB1_HSV6U
19	94.5	3.8	1132	1	YABD_SCHPO
20	94.5	3.8	1616	1	UBR1_SCHPO
21	93.5	3.8	1958	1	P1P_LACIC
22	92.5	3.7	1902	1	VGLE_MOUSE
23	92	3.7	538	1	NPH1_MOUSE
24	91	3.7	253	1	SCAA_RAT
25	91	3.7	698	1	MUTS_BUCAT
26	91	3.7	802	1	DNB1_HSV6U
27	91	3.7	1132	1	CACP_CANTR
28	90.5	3.6	627	1	MYSH_LACAC
29	90.5	3.6	1577	1	P2P_LACIA
30	90	3.6	1902	1	VGLE_MOUSE
31	90	3.6	538	1	MUTL_BORBU
32	90	3.6	610	1	
33	90	3.6	610	1	

34	89.5	3.6	569	1	V139_MYCCE
35	89.5	3.6	668	1	PSB2_YEAST
36	89.5	3.6	1108	1	NAH3_DIDMA
37	89.5	3.6	839	1	CN3B_RAT
38	89.5	3.6	1857	1	FAS2_PENPA
39	89.5	3.6	382	1	VMAT_SVAL
40	89	3.6	538	1	VGLE_MOUSE
41	89	3.6	699	1	DPOI_MOUSE
42	89	3.6	908	1	DPOI_MOUSE
43	89	3.6	959	1	TELI_YEAST
44	89	3.6	2787	1	TELI_YEAST
45	89.5	3.6	230	1	HBP_MAIZE

ALIGNMENTS

RESULT 1
ID BRL1_RABIT STANDARD: PRT: 540 AA.

AC 005004: BRL1_RABIT STANDARD: PRT: 540 AA.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE BRUSH BORDER 61.9 KDA PROTEIN PRECURSOR.
GN ADRAB-A.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA MEDLINE=93286138; PubMed=8509424;
RA Boll W., Schmid-Chanda T., Semenza G., Mantel N.;
RT "Messenger RNAs expressed in intestine of adult but not baby rabbits.
RT Isolation of cognate cDNAs and characterization of a novel brush
RT border protein with esterase and phospholipase activity.";
RJ J. Biol. Chem. 268:12901-12911(1993).
CC -1- TISSUE SPECIFICITY: INTESTINE, AND IN A LESSER EXTENT IN KIDNEY.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE INTESTINE OF ADULT BUT NOT
CC BABY RABBIT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: 212840; CAY8302.1; -
DR PIR: B45665; B45665.
KW SIGNAL.
FT CHAIN 1 22 POTENTIAL.
FT CHAIN 23 540 BRUSH BORDER 61.9 KDA PROTEIN.
SQ SEQUENCE 540 AA: 61888 MW: EB6570EC86925A CRC64;

Query Match 64.6%; Score 1606; DB 1; Length 540;
Best Local Similarity 57.3%; Pred. No. 6.3e-17;
Matches 309; Conservative 67; Mismatches 91; Indels 72; Gaps 2;

OY 8 YKSLALLFLILASIIIFVFNQSKVSAISLSLHWNSTKSLFPTPLISKPLTE 67
DB 3 HKYKLLCLLAACVLCIISONSSTKIGALKPNHYSNMSISSIPKMSVPKSLTE 62
OY 68 TELKLEIKLEKLDQOIPRPPTHVNTTSATSTATILNPDVYCGDOLHILEVDHL 127
DB 63 TELKLEIKLEKLDRIIPRPPTHVNTTSATSTATILNPDVYCGDOLHILEVDYL 122
OY 128 GRROKYGSEFLARRSKSPALMAAGSKVTDENNGTYLVSEFLWEGOVSLILLHPSEG 187

Db 123 CHQKEVGGDFLRARFSPALAGSGVTDENNNYIVLFTLMEQVLSVLLIHSEG 182
 QY 188 VSALMSARNOGYDRIYFTGQVNGTSGVHSEGLLNTNELCOYLDNDQDGFYVAPQ 247
 183 ASALMRARNOGYDRIYFTGQVNGTSHVTEGSLNTSNECKYLNGRDQVFECKMPQ 242
 QY 248 HMPCAALTHMYSNKKVSYLSKQKSLFERSVNGVEIMKENTISYSKN----- 297
 Db 243 HMPCEALTHYSKNDISYLSKSKNLFHRSKVGEIM-KKHQIDVSCNKSKEYEKCQ 301
 QY 298 ----- 297
 Db 302 IGKIPVPGGYTLGGRLTFPCNQIQDPAKISGCKGLIYLMGDSLRQWYILPKVM 361
 QY 298 -TLKSVLHESGKLQHLAVDLDNINIMOKYCPILGSMYSVEMMYLRAIDRPG 356
 362 KTLKFPELHETGKFKHLLDAEKNHQLQKNSHPFTVQLSVSDHGYIPQEDRLIG 421
 QY 357 EKNTIVISLGHFRFPIDVFIRALNKAIOHLLRSPDMVITKENTREMYNDAAE 416
 422 DKDTIVITFGHFHFPIDIFIRASVQAIEFLRSPATKVIKTEINREMHIEAE 481
 Db 417 RPSDFGYIYGLIKDFIDSLDSDVSIIDMDITATGCTNNHPPQHVGNQIMLLNTYC 475
 482 RFGDFHYIYGLIKDFIKDLNVGVNDAMDNTIATGNNHPPQVYIGNIMFLNTYC 540
 Db
 RESULT 2
 ID APB_HUMAN STANDARD: PRT: 4563 AA.
 AC P0414; 000502; Q13787;
 DT 01-NOV-1986 (Rel. 03. Created)
 DT 01-NOV-1986 (Rel. 03. Last sequence update)
 DT 01-NOV-2000 (Rel. 40. Last annotation update)
 DE APOLIPOPROTEIN B-100 PRECURSOR (APO B-100) [CONTAINS: APOLIPOPROTEIN
 DE B-48 (APO B-48)].
 GN APOB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87016385; PubMed=3763409;
 Kuoct.T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J.,
 Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
 *Complete cDNA and derived protein sequence of human apolipoprotein
 B-100.*
 RT Proc. Natl. Acad. Sci. U.S.A. 83:7501-7503(1986).
 RN [12]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88003974; PubMed=3652907;
 Ludwig E.H., Blackhart B.D., Pterotti V.R., Calati L., Fortier C.,
 Kuoct.T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;
 DNA sequence of the human apolipoprotein B gene.
 RT DNA 6:363-372(1987).
 RN [13]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87008488; PubMed=3759943;
 Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H.,
 Gotto A.M., Jr., Chan L.;
 *The complete cDNA and amino acid sequence of human apolipoprotein
 B-100.*
 RT J. Biol. Chem. 261:12918-12921(1986).
 RN [41]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87041416; PubMed=346946;
 Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J.,
 Lee N., Brewer H.B., Jr.;
 *Human liver apolipoprotein B-100 cDNA: complete nucleic acid and
 RT derived amino acid sequence.*
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87161758; PubMed=3030729;
 RA Cladarias C., Hadzopoulou-Cladarias M., Nolte K.T., Atkinson C.,
 RA Zannis V.I.;
 *The complete sequence and structural analysis of human
 RT apolipoprotein B-100: relationship between apob-100 and apob-48
 RT forms.*
 RL EMBO J. 5:3495-3507(1986).
 RN [6]
 RP SEQUENCE OF 709-906 FROM N.A.
 RX MEDLINE=85270450; PubMed=3860836;
 RA Deeb S.S., Motulsky A.G., Albers J.J.;
 A partial cDNA clone for human apolipoprotein H.
 RT Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
 RN [17]
 RP SEQUENCE OF 3056-3159 FROM N.A.
 RX MEDLINE=86041888; PubMed=3903660;
 RA Mehrabian M., Schumaker V.N., Fareed G.C., West R., Johnson D.F.,
 RA Kirchgesner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.;
 *Human apolipoprotein B: identification of cDNA clones and
 RT characterization of mRNA.*
 RL Nucleic Acids Res. 13:6937-6953(1985).
 RN [18]
 RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.
 RX MEDLINE=86093680; PubMed=3841204;
 RA Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O.,
 RA Bjursell G.;
 Molecular cloning of human apolipoprotein B cDNA.
 RL Nucleic Acids Res. 13:8813-8826(1985).
 RN [19]
 RP SEQUENCE OF 3109-4563 FROM N.A.
 RX MEDLINE=85300528; PubMed=29934225;
 RA Kuoct T.J., Rall S.C., Jr., Innerarity T.L., Jacobson S.F.,
 RA Udeal M.S., Levy-Wilson B., Powell L.M., Pease R.J., Edy R.,
 RA Nakai H., Byers M., Priestley L.M., Robertson E., Rall L.B.,
 RA Besholtz C., Shows T.B., Mahley R.W., Scott J.;
 *Human apolipoprotein B: structure of carboxyl-terminal domains,
 RT sites of gene expression, and chromosomal localization.*
 RL Science 230:37-43(1985).
 RN [10]
 RP SEQUENCE OF 1-291 FROM N.A.
 RX MEDLINE=86149325; PubMed=3513177;
 RA Procter A.A., Hardman D.A., Schilling J.M., Miller J., Appleby V.,
 RA Chen G.C., Kirsch S.W., McIntire G., Kane J.P.;
 *Isolation of a cDNA clone encoding the amino-terminal portion of
 RT human apolipoprotein B.*
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
 RN [11]
 RP SEQUENCE OF 1-1670 FROM N.A.
 RX MEDLINE=86287319; PubMed=3461454;
 RA Procter A.A., Hardman D.A., Sato K.Y., Schilling J.M.,
 RA Yamanaka M., Hori Y.J., Hyerrillid K.A., Chen G.C., Kane J.P.;
 *Analysis of cDNA clones encoding the entire B-26 region of human
 RT apolipoprotein B.*
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
 RN [12]
 RP PARTIAL SEQUENCE, AND IDENTIFICATION (APO-B48).
 RX MEDLINE=88018019; PubMed=3659919;
 RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.-P., Weng S.-H.,
 RA Silberman S.R., Cai S.-J., Deslypere J.P., Rosseneu M.,
 RA Gotto A.M., Jr., Li W.-H., Chan L.;
 *Apolipoprotein B-48 is the product of a messenger RNA with an organ-
 RT specific in-frame stop codon.*
 RL Science 238:363-366(1987).
 RN [13]
 RP DOMAINS.
 RX MEDLINE=87039351; PubMed=3773997;
 RA Kuoct T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C., Jr.,
 RA Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R.,
 RA Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W.,
 RA Levy-Wilson B., Scott J.;
 *Complete protein sequence and identification of structural domains
 RT of human apolipoprotein B.*

Fri Oct 12 16:09:16 2001

us-09-729-454-1.ra1

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: October 12, 2001, 16:02:37 ; Search time 20.85 Seconds

(without alignments)
469,085 Million cell updates/sec

Title: US-09-729-454-1
Perfect score: 2485

Sequence: 1 MKISMINKYKSLALLFILAS.....VHPQHVGNQINILNLATIC 475

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 segs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents: AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/6D.COMB.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	104	4.2	1494	3	US-08-755-587-186
2	100	4.0	1007	3	US-08-851-843A-86
3	100	4.0	1007	4	US-08-974-549A-187
4	100	4.0	1031	3	US-08-851-843A-2
5	100	4.0	1031	4	US-08-974-549A-110
6	95	3.8	610	2	US-08-879-561-10
7	94.5	3.8	1579	3	US-08-755-587-184
8	92	3.7	849	3	US-08-804-439A-17
9	92	3.7	849	3	US-08-720-229-17
10	90.5	3.6	2325	4	US-08-417-089-6
11	90.5	3.6	2325	4	US-08-695-651-6
12	90.5	3.6	2325	4	US-08-930-285-6
13	90	3.6	668	1	US-08-376-362A-20
14	89.5	3.6	668	1	US-08-530-950-13
15	89.5	3.6	668	4	US-09-149-879-13
16	87.5	3.5	315	1	US-08-571-758-12
17	87.5	3.5	315	1	US-08-909-984A-12
18	87.5	3.5	315	1	US-08-909-983-12
19	87.5	3.5	648	1	US-08-185-282-4
20	87.5	3.5	3418	2	US-08-639-501-2
21	87.5	3.5	3418	3	US-09-044-946-2
22	87.5	3.5	3418	3	US-09-044-908-2
23	86.5	3.5	648	1	US-08-185-282-3
24	86.5	3.5	2329	3	US-08-755-587-16
25	86	3.5	395	1	US-07-931-943-5
26	85.5	3.4	346	1	US-08-276-151-5
27	85.5	3.4	648	1	US-08-276-151-2

28	85.5	3.4	648	1	US-08-185-282-12	Sequence 12, Appl
29	85.5	3.4	648	2	US-08-886-751A-6	Sequence 6, Appl1
30	85.5	3.4	648	3	US-09-209-668-13	Sequence 13, Appl1
31	85.5	3.4	3418	2	US-08-603-753D-4	Sequence 4, Appl1
32	85.5	3.4	3418	3	US-08-755-587-44	Sequence 44, Appl1
33	85.5	3.4	3418	4	US-09-099-753-4	Sequence 4, Appl1
34	85.5	3.4	3418	4	US-08-986-106-4	Sequence 4, Appl1
35	84.5	3.4	387	4	US-09-323-427-4	Sequence 4, Appl1
36	84.5	3.4	521	2	US-08-737-825-3	Sequence 3, Appl1
37	84.5	3.4	648	1	US-08-185-282-1	Sequence 1, Appl1
38	84.5	3.4	648	1	US-08-185-282-2	Sequence 2, Appl1
39	84	3.4	417	1	US-08-351-981-7	Sequence 7, Appl1
40	83	3.3	395	1	US-07-931-943-2	Sequence 2, Appl1
41	83	3.3	395	2	US-08-624-601-15	Sequence 15, Appl1
42	83	3.3	417	1	US-08-351-981-6	Sequence 6, Appl1
43	83	3.3	831	2	US-08-677-734A-11	Sequence 11, Appl1
44	83	3.3	874	3	US-08-804-439A-15	Sequence 15, Appl1
45	83	3.3	874	3	US-08-720-229-15	Sequence 15, Appl1

ALIGNMENTS

RESULT 1
US-08-755-587-186
Sequence 186, Application US/08755587
Patent No. 604597
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
TITLE OF INVENTION: susceptibility gene and uses thereof.
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 186:
SEQUENCE CHARACTERISTICS:
LENGTH: 1494 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-755-587-186
Query Match 4.2%, Score 104; DB 3; Length 1494;
Best Local Similarity 21.3%; Pred. No. 0.3;

Matches 79; Conservative 54; Mismatches 142; Indels 96; Gaps 19;

```

QY 22 IIFVONSTKYWSALNLSLHYMNNSTKSLFKPTLISLK-----PLTELEL- 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 IISGFVNSTY-----SDESGHTAPPTLSLKQDDSNRNTTPSQKAEIT 477
QY 71 RIKEIIEKLDQIIPRPF---TW--NTTSATHTATILNPDTYCRGDLHLLEVD 125
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 478 ELSTILEESGSOFEFTQFRKPSHILOKNPFMPENQTLILNSTKEMWDDHLTTAPS 537
QY 126 -HLGRRKQ---YGGDFLRARMSPALMAGAGKYTDNNQTLVSEFLFMGQ-VSLSL 180
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 538 IQVDSKSSGILIGKOKFACLSLSTNSASGYSTDKN-----VEFGFYASRQTKLV- 592
QY 181 LHPSEGVSAIWSAR-----NGYDRVLTGQFVNGTSQV----- 215
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 593 -----GSEALQAKKLLFSDLENINETSVEVDRSFSSSYNDVSVMIOIDCNKNNEP 646
QY 216 HSECGILINTNAELC-----QYLDNRDQGFYCVRPQHMCALTL--HMYSKNK 263
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 647 NNKCRLLIGNNIEMTIDIFVEETESYRNRTEENGNOCTDAGRNTKNSGSDGSKNDT 706
QY 264 VSYLSKQESL--FERNNGVEIMKEF---NT-----ISVSKNTLKSVDL-HESGKLQ 311
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 707 V-TIHEENGLPCTIDQHNIDLKLFSQFMKEGNTQIKESLSDTLCLVMKAEETSHVTMSK 765
QY 312 HOLAVDLDRNT 322
    ||| : |||
Db 766 QQLTANTGONI 776

```

RESULT 2

```

; Sequence 86, Application US/08851843A
; Patent No. 6093809
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Hailey, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6093809e1 Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,843A
; FILING DATE: 06-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION:

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015369-0029300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0300
; TELEFAX: (415) 576-0200
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1007 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-851-843A-86

```

Query Match 4.0%; Score 100; Db 3; Length 1007;
Best Local Similarity 21.5%; Pred. No. 0.4;
Matches 106; Conservative 71; Mismatches 179; Indels 136; Gaps 27

```

QY 18 LASWIIIFVONSTKYWSALNLSLHYMNNSTKSLFKPTLISL-KILHTELK- 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29 LYSWT-----QKVICRNSOS-HKLEDEIKIFAOTNIVATPRVHEEFKVIYANK 79
QY 74 EI-----IEKLDQIIPRPFTVNTTSATHTATILNPDTYCRGDL- 124
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 80 EYFSTGLMELIDKCL-----VELSSSVSDRQKLCQGFQIKGQVLAKTHLTA- 130
QY 125 DHLGRKQYGDFLRARMSSPALMAGAGKYTDNNQTLVSEFLFMGQVSLIHP 184
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 131 --LSTOKY--FFQDENNVAMIG-----NELFRHLTKY-----LIFORT 186
QY 185 SEGYSALMSARNGYDRVLTGQF---VNGTSQVH-----SECGILINTNAL-VYLDN 235
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 169 SEGTLVQFCGNV-FDHLKVNDKDPKQKGAADNMEPPCSTCK--YNNKNDKQHLNN 245
QY 236 KDQEG-----FYCVRPQHMCALTHMYSKNKQSYLSQKESLFEERNVVEIM 265
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 226 INVNNMNMKSRTIRYC-----THF--NRNNQFFKKHE--FVSKNNISGM 268
QY 286 EKFNIT--SVSKNTL-KSYDLHESGKLQGLAVDLDRKINIQOKQCYTLJSMTYSVK 344
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 265 DRAQITFTNIFNRIRKKLQKQVIEKIAVMEKRYDNPVYLTQSC-PLPENMKRQ 327
QY 343 EMEVLTIRADTSGEK-----NTVVISLQGH-RPPIDVFIR----- 381
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 328 KIENL--INKTRESKSYEELFSYTDNKKCYOTQINPEFYNILLPDLTGNRKNQK 384
QY 382 -----ALNVKAI-OHLIRSPTMTY----IKTENIREMYNDAERFSFHGYIYLI 430
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 385 KKKVVELNKHHLIRKHLLEKINTREISMVQVEISAKHYTFDHEMT-----YVLMKEL 439
QY 431 KDIFQDLSVIT 442
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 440 RWIPEDLVSLI 451

```

RESULT 3

```

; Sequence 187, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Hailey, Calvin
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:

```


Fri Oct 12 16:09:16 2001

us-09-729-454-1.rai

Page 3

```

ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-00261005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ. ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 1007 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-974-549A-187

Query Match 4.0%; Score 100; DB 4; Length 1007;
Best Local Similarity 21.5%; Pred. No. 0.4;
Matches 106; Conservative 71; Mismatches 179; Indels 136; Gaps 29;
UY 18 LASMTITFVONSTKWSALNLSIAHWNNSKSLPPTPLISL-KPLETELEIRI---K 73
DB 29 LKSWI-----OKVIRCRNOSOS-HYKDELEIKIFAQTNIVATPRDYNEDEPKYIARK 79
OY 74 EI-----IEKLDQOIPPPFTHVNTTSATSHSTATILNPRDYCGRGQUL---HILLEVR 124
DB 80 EVFSTGLMIELEIKCL-----VELLSSSLVSDRQKIDCGFQULKGNQIAKTHLTLTA-- 130
```

```

OY 125 DHGRRROYGSDFLBARSSPALMAGASKYTFDNNGTYLVSTFLFNEGVSLSLLIHP 184
DB 131 --LSTOKOY---FQDEMNNOVAMIG-----NELFPHLYTRY-----LIFORT 168
OY 185 SEGVSAIWSARNQGYDRYIFTGF---VNGTSQVH-----SECGLIINTNAELCOYLDN 235
DB 169 SEGTLVQFCGNNV-FDHLKYNDKDFDKQKGAADNBPCCSTCK--YNNKNDHPLNN 225
OY 236 RDEEG-----FYCVRPQHPICALTHWYSKNKVSYSLSKQKSLPERSVNGVEIM 285
DB 226 INPNNNNNNKSRIRFLFC-----THF---NNANNOFRKHE---FVSNNNNSGM 268
OY 286 EKNTT--SVSKCNTL-KSYDLHESGKIQHQLAVDLDRNINIQOWKYCPPLIGSMYSVK 342
DB 269 DRAGTFTNIFEFNRIRKRLKDKVIEKIAMYLEKVDNFENYLTFRSC-PLPERWPERQ 327
OY 343 EMEYLFRAIDRTGGEK-----NTVIYISIQHF-RPPIDVFIIR----- 381
DB 328 KIENL---INTREKSKYEELEFSTTDNKCVCYQFINFEFTNILLKDFLTGRNRKNFK 384
OY 382 -----ALNVKAI-QHLLRSPDTWVI---IKTENIREMYNDARESDPHGYOYLII 430
DB 385 KVKKYVELNKHLELHKMLLEKINTREISMNQVETSAKHRYFDHENDI-----YVLKML 439
OY 431 KDIFODLSYSII 442
DB 440 RMIFEDLVYSII 451

RESULT 4
US-08-851-843A-2
Sequence 2, Application US/08851843A
Patent No. 6093809
GENERAL INFORMATION:
APPLICANT: Geoh, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6093809e1 Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,843A
FILING DATE: 06-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
```

NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1031 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 US-08-851-843A-2

Query Match 4.0%; Score 100; DB 3; Length 1031;
 Best Local Similarity 21.5%; Pred. No. 0.41;
 Matches 106; Conservative 71; Mismatches 179; Indels 136; Gaps 29;

QY 18 LASMIITVFONSTKWSALNLSLHYNNSTKSLPPTPLSL-KPLETELEK-
 Db 30 LYSWI-----QKVICRNSOS-HKDELDKIFAOINIVATPRDYNEEDFKVIARK 80
 QY 74 EI-----IEKLDQIIPPRFTHVNTTTSATHTATILNPRDYCRGDQ-
 Db 81 EVFSTGLMIELDKL-----VELSSSVSDROKLCFCGFGOLKGNOLAKTHLLTA-- 131
 QY 125 DHLGRKQYGGDFLRAMSSPALMAGAGKVTDFNNGTYLVSFTLFWEQVSLILLHP 184
 Db 132 --LSTOKY--FFQEDMNVAMIG-----NELFRHLYTKY-----LIFORT 169
 QY 185 SEGVSALMSARNOGYDRAVFTQF-----VNGTSQVH-----SEGLILNTMAELCOYLDN 235
 Db 170 SEGTLVQFCGNV-FDHLVNDKFDKQKGGADNNEPRCSCTK--YANKNKHQFLNN 226
 QY 236 KDQEG-----FYCVRPQHPICALTHMYSKKKYSYLSQKESLPERSNVGEIM 285
 Db 227 INVPMNMNKSRTRIEYC-----THF--NRNNGFFKHE--FVSKNNISAM 269
 QY 286 EKFNIT--SVSKCNTL-KSYDLHESGKLOQLAVDLDRININIOKCYPLIGSMYSVK 342
 Db 270 DDAOTIFTNIFRNRIKKLADKVIETIAVMEKYKDFNFVYLTKSC-PLPENRERKQ 328
 QY 343 EMEYILTRADRTGGEK-----NTVVISLQOH-REPPIDVEIRR----- 381
 Db 329 KIENL---INKTREEKSKYYEELFSYTDNKCQVQFINEFYNILPRDELTRNKNKQK 385
 QY 382 -----ALNVHKA-CHLLRSPTWYI---IKTENIREKYNDAERFSDHYQYI 430
 Db 386 KVKKYVELNKHLEIKHKNLLEKINTREISMQVETSAKHFYEDHENT---VYLKRL 440
 QY 431 KDFQDLVSII 442
 Db 441 RNIFEDLWVSLI 452

RESULT 5
 US-08-974-549A-110
 ; Sequence 110, Application US/08974549A
 ; Patent No. 6166178

GENERAL INFORMATION:
 APPLICANT: Cecn, Thomas R.
 APPLICANT: Lingner, Joachim
 APPLICANT: Nakamura, Toru
 APPLICANT: Chapman, Karen B.
 APPLICANT: Morlin, Gregg B.
 APPLICANT: Harley, Calvin B.
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: Human Telomerase Catalytic Subunit
 NUMBER OF SEQUENCES: 727
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/974,549A
 FILING DATE: 19-NOV-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/851,843
 FILING DATE: 06-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/854,050
 FILING DATE: 09-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/911,312
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/912,951
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/915,503
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/17618
 FILING DATE: 01-OCT-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/17885
 FILING DATE: 01-OCT-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph Ted
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 110:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1031 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-974-549A-110

Query Match 4.0%; Score 100; DB 4; Length 1031;
 Best Local Similarity 21.5%; Pred. No. 0.41;
 Matches 106; Conservative 71; Mismatches 179; Indels 136; Gaps 29;

QY 18 LASMIITVFONSTKWSALNLSLHYNNSTKSLPPTPLSL-KPLETELEK-
 Db 30 LYSWI-----QKVICRNSOS-HKDELDKIFAOINIVATPRDYNEEDFKVIARK 80
 QY 74 EI-----IEKLDQIIPPRFTHVNTTTSATHTATILNPRDYCRGDQ-
 Db 81 EVFSTGLMIELDKL-----VELSSSVSDROKLCFCGFGOLKGNOLAKTHLLTA-- 131
 QY 125 DHLGRKQYGGDFLRAMSSPALMAGAGKVTDFNNGTYLVSFTLFWEQVSLILLHP 184

Query Match 3.8%; Score 94.5; DB 3; Length 1579;
 Best Local Similarity 19.6%; Pred. No. 3.1; Mismatches 169; Indels 181; Gaps 23;
 Matches 102; Conservative 69;

QY 48 NSTKSLFP--KTPPLSLKPLTE-----TELRIKE-IIEKLDQIIPRPETHVNTTS 96
 DB 431 NSNNLTPSOAKETELSTLESQSEFFTOFRKPSYIIQKSTFEVPENMOTILKTTSF 490
 QY 97 ATHSTATILNPDYTCRGDQLHILLEVD-HIGRRKQYGGDFLRAMSGPALM---AGA 151
 DB 491 E-----CRDADLHYIMNAPSIOYDSSKQFECTVEIKRKAGLLKNDCKSA 536
 QY 152 SGKVTDFN-----NGTIYVFTLFMEGOVSLSLLIHPSGVSA---LMSARNO 197
 DB 537 SGYLTDENEVGRFGYSNHTKLNSTELQAVKLFSDIENISETSAEVHPISLSSSK 596
 QY 198 GYDRVITFGQFVN---GTSQVHSECGLLIINTNAELC-----OYLDNRQDE----- 239
 DB 597 CHDSVSMFKIENHNNDKTSKNNKCOLLONNIMETGTVEITENKRTENEDNKYT 656
 QY 240 -----GFYGVYR-----PQHPGALTHMYSK----- 260
 DB 657 AASRNHNLEFDGSDSKNDYVCIHKDETDLFTDOHNTCLKLSQGFKEGNTQIKEDLS 716
 QY 261 -----NKKVSYLSKQESL--FERSNVGVEIMEKFNITSVSKCNTLK 300
 DB 717 DLTFLVAKAOEACHGNTSNKEQLTATEQNINKDFSTDPFQTASGKN-ISVAKSEFK 775
 QY 301 SV-----DLHESGQLQHOLAVIDLN-----IN 323
 DB 776 IVNFEOKPEELH-NFSLNSELHSDIRKNNKMDILSEETDIYKHLKESPVGTGQDLVT 834
 QY 324 IOWO-----KYCYP-LIGSWTYSVAKEMEYLRADRTG-----EKNTVIYISLQGHFPP 372
 DB 835 FQGOERDEKIKEPTLLGFTASGKKVKIKESLIDVKMLFDEKQGGPSEITSFSGNW-- 892
 QY 373 FPIDFIRRALNVHKAIOHLLRSPDPTVIKTENIREMYN 413
 DB 893 -----AKTLKYREACKDEL-AGETIETIAPACKCKEMON 925

RESULT 8
 US-08-804-439A-17
 ; Sequence 17, Application US/08804439A
 ; Patent No. 6015565
 ; GENERAL INFORMATION:
 ; APPLICANT: Rose, Timothy M.
 ; APPLICANT: Bosch, Marlin L.
 ; APPLICANT: Strand, Kurt
 ; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV
 ; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES
 ; NUMBER OF SEQUENCES: 113
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 4225 Executive Square, Ste 1400
 ; CITY: La Jolla
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: February 21, 1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hallie, Lisa A.
 ; REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 09176/004001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 678-5070
 TELEFAX: (619) 678-5099
 TELEX:
 ; INFORMATION FOR SEQ ID NO: 17:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 849 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-804-439A-17

Query Match 3.7%; Score 92; DB 3; Length 849;
 Best Local Similarity 18.7%; Pred. No. 2;
 Matches 86; Conservative 82; Mismatches 172; Indels 120; Gaps 20;

QY 25 TYFQNSTVWSA---LNLIS--LHYNN-----STKSLFPKP--LISKPLTEELR 71
 DB 353 TEVENFTSTYSCVSDQINKTISEYIOLKNSYVAGSKTYQFTDGNLYLIMQPLEHPE-- 410
 QY 72 IKEITEKLDQIIPRPETHVNTTSATSTATILNPDYTCRGDQLHILLEVRDHLGRKK 131
 DB 411 -----IEDIDEDSDPEPTPAPPKSTRKRREADNGNSTSEVSGSENPILT-----AOI 459
 QY 132 OYGGDFELARRSS---PALMAGASGYTFDFNGIYVSLTFMEGOVSLSLLIHPSGV 188
 DB 460 QPAYDKLITVSNVLELSRAMCRQVND-----TLKW-----YELSVNNTSVM 504
 QY 189 SALW-----SARNG-----YRNVITQYVNGIS 213
 DB 505 SAIYKPKPARYVGDIASTDCITYDQSSVNIHOSLRLOHDKTTYSNPFVYTKFINST 564
 QY 214 QVHSCG-----LILTNALCOYLDRNDGEFCYRPUHPCALTHMYSKKKVSYLS 268
 DB 565 PLTGQGRREILINSTNTEG-----KDESEHYFVGEYI-----YYKN---YIF 608
 QY 269 KQESLFEERSNVGVEIMEKFNITSVSKCNTL--KSYDLHESGQLQHOLAVIDLNINQW 726
 DB 609 EKKLNL-----SSIALDLOFTALINSTFIENIDKTYELVSSTERKLASSVFOIESREY 663
 QY 327 QKYCYPILG-----SMTYS---VKEMEYLRADRTGSEKNTVIYISLQGHFPPYI 375
 DB 664 NYTYSLAGIKKDLDTIDYNDRLYODLSDMAADLGIG--RSVNVWSSVYTFESSIV 721
 QY 376 DVFIRRALNVHKAIOHLLRSPDPTVIKTENIREMYN 415
 DB 722 TGFIKFTNPLGCTIFILLIGITFLVVVILNRNSQFIDA 761

RESULT 9
 US-08-720-229-17
 ; Sequence 17, Application US/08720229
 ; Patent No. 6022542
 ; GENERAL INFORMATION:
 ; APPLICANT: Rose, Timothy M.
 ; APPLICANT: Bosch, Marlin L.
 ; APPLICANT: Strand, Kurt
 ; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV
 ; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES
 ; NUMBER OF SEQUENCES: 100
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Morrison & Foerster
 ; STREET: 755 Page Mill Road
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304-1018
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/720,229
: FILING DATE: 26-SEP-1996
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Schiff, J. Michael
: REGISTRATION NUMBER: 40,253
: REFERENCE/DOCKET NUMBER: 29938-20002.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 813-5600
: TELEFAX: (415) 494-0792
: TELEX: 706141
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 849 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-720-229-17

```

```

Query Match      3.7%; Score 92; DB 3; Length 849;
Best Local Similarity 18.7%; Pred. No. 2;
Matches 86; Conservative 82; Mismatches 172; Indels 120; Gaps 20;

```

```

QY 25 TWFNSTKVMAS---LNLIS--LHYNN---STKSLFPTP---DISKPLTELELR 71
N 35 TEVNFSTIYSCVSDQINKTISBYLOKINSYVASKGTQYFKTDGNLYLMOLENE-- 410
QY 72 IKEIIEKLDQIIPRPTHVNTTTSATSHSTATILNPRDTYCRDGLHILLEVDHGLRRK 131
Db 411 ----TEDIDEDSOEPTPPAPKSTRKREADNCSNTEVSYSKSENPLIT-----AQI 459
QY 142 QYGDPLFARMSS---PALMAGASGKVTDFNNGTIVLSFTLFMEGOVSLSLIHSEGV 188
Db 460 QFAYDKLTISVNVYLFELSMACREQVARD-----TLWM-----YELSKNPTISVA 504
QY 149 SALM-----SARNQ-----YDVNITGOVFNVS 213
Db 605 SNIYGRVMAAYVGDIASTDCIYVDSSVNIHQSLRLOHDKTTCYSRPRVTEFFINSTD 564
QY 214 OVHSECG-----LILTNALCOYLDNRDQGFYCVARPOHPCALTHMYSKKNKYSYL 268
Db 565 PLTGQIGPRKEIILSNNIETC-----KDESEHYETGEY-----YKKN-----YIF 608
QY 269 KOEKSLEFRSNVGEIEMKFNISVSKCNLT--KSYDLHESGKLOHOLANDLRNINIQM 326
Db 609 EELNLT-----SIALDIFIALNISFTLENDRTVELYSTERKLASSVFDIESMEREX 663
QY 327 QKYCYPLIG-----SMYS---VKEMEYLTRAIDRTGGEKNTVYISLGHFFRPPFI 375
Db 664 NYTYSLAGIKKDLNDTIDYNRDLVODLSDMADLDGIG--RSVNVVSSVYTFPSSIV 721
QY 376 DVFIRRALNVHKAIOHLRLSPDTWIIKTENIREMYNDA 415
Db 722 TGFIKFTNPLGGITILLIGGIIFLVVYLNRRNSQFHDA 761

```

```

RESULT 10
US-08-417-089-6
: Sequence 6, Application US/08417089
: Patent No. 6069298
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: METHODS AND AN ACETYL COA CARBOXYLASE GENE
: TITLE OF INVENTION: FOR CONFERRING HERBICIDE TOLERANCE AND AN ALTERATION IN
: NUMBER OF SEQUENCES: 11
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

```

: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/417,089
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2325 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-417-089-6

```

```

Query Match      3.6%; Score 90.5; DB 3; Length 2325;
Best Local Similarity 18.8%; Pred. No. 16;
Matches 99; Conservative 78; Mismatches 178; Indels 171; Gaps 23;

```

```

QY 63 KPLTELEIKETIEKLDQ-QIPRPTHVNTTTS-----ATHSTATILNPRDTY----- 111
Db 659 KVTYTNAAVSEYVSYLTGKHIPKHISIVNSTVNIIGSKYLTETRTGHSYRLRBN 718
QY 112 -----CRGDQ-----HILLEVDHGRKQYGGD--FLRARMSPALMAGA 151
Db 719 DSTVEANVSLDGLMOLDGNSHVIAEEAGNTRLDQDKTCLLDNHDPSKLLAET 778
QY 152 SGAVYDF--NNGTIVSFTLFMEGOV--SLSLIHSEGOVSLMSARNQGT---DVT-- 203
Db 779 PCKLRLFLVADGADVADYAEVEMKCMKPLLSVPSAGVYIHWMSSEGQALQADLIARL 838
QY 204 -----FTGQF-----VNGTSQVHSECGLIINT----- 225
Db 839 DLDDSAVKAAREPDGIFPOMELPVAVSSQVHKRYAASLNAARVYLAGEHNINEVVDL 898
QY 226 -----NAFL--COY-----LDNRDQ--GFYCVARPOHPCALTHM 257
Db 899 VCCIDNPEPLPLQNDLMSYLATRLPRLNKLSELDQYKYNLFGHGNDFPSKLLRDI 958
QY 258 YSKNNKVSYSKQKSLFER-----SNVGEIEMKFNISVSKCN 297
Db 959 IEBN--LSYGSEKKEKATNERLVEPLMLNLKSYEGRESHAFHVVKSLEFEYLVELPSD 1016
QY 298 TLKSYDLHESGKLOHOLANDLRNINIQMOCYCPPLIGSMYSVKEMEYLTRAIDTQGE 357
Db 1017 GLOS-DVLET--LRHQSKDQKQVDT-----VLSHQVYR 1048
QY 358 KNTIVISLGQH--RFP--IDVFIRRALNVHKAIOHLRLSPDTWIIKTENIREMYN 413
Db 1049 NNAKLVTALMEKLYVNPQGYRDLVFRSSLNHKRYKALAKASLELQTKSELRA--S 1106
QY 414 DAERFSDPHGYOYLIFKIDIFQDL-----SVSIIDAMDITI 449
Db 1107 VARSLSDLGMHKGEMSKDMEDLYSAPLPVEDALISLPDYSDRIV 1152

```

```

RESULT 11
US-08-695-651-6
: Sequence 6, Application US/08695651
: Patent No. 6146867
: GENERAL INFORMATION:
: APPLICANT: Geogenbach, B. G.
: APPLICANT: Somers, D. A.
: APPLICANT: Wyse, D. L.
: APPLICANT: Gronwald, J. W.
: APPLICANT: Egli, M. A.
: APPLICANT: Lutz, S. M.
: TITLE OF INVENTION: METHOD FOR PRODUCING PLANT ACETYL COA
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
: STREET: P.O. Box 2938
: CITY: Minneapolis

```



```

QY 204 -----FTGOF-----VNGTSOVHSECCLLNT----- 225
Db 839 DLDPSAVKRAEPFDGIFPOMELPRAVSSOVHKKYVAAASLMAARMLAGYEININEVYODL 898
QY 226 -----NAEL--COY-----LDNRDQ-----GFYCVRPHMPCALTHM 257
Db 899 VCCLDNPELPLOMDLSVYATRLPRNLKSELEDKYEKIKAFYHCKNEDFPKLLROI 958
QY 258 YSKNKKVSYLSKOKSLPER-----SNGVEIMEKNTSVSKCN 297
Db 959 IEE--LTSYSEKEMKATNERLEVERPLMLNKSEEGNEAHNVVSLSEELIVEELFSD 1016
QY 298 TLKSYDLHESGKLOHOLAVLDLRNINIQMOKYCPILGSMYTSVEMEXYLRALDRTGE 357
Db 1017 GLOS-DVLET--LRHOSKLOKQVYDI-----VLSHQVVR 1048
QY 358 KNTVIVISLQOHF-RFP--IDVFIRRALNVHKAIOHLLKSPDTWYIKTENIREMN 413
Db 1049 NKAKLVYALMEKLVYVPPGKDLVFRSSLNKRYKALALASLEEQTKSELRA--S 1106
QY 414 DAEFSDPHGYIOYLKIDFODL-----SVSIIDANDTTI 449
Db 1107 VARSLDGMHKGEMSIKDMEDVSAFLPYEDALISLFDYSDRTV 1152

RESULT 13
US-08-376-362A-20
; Sequence 20, Application US/08376362A
; Patent No. 5693756
; GENERAL INFORMATION:
; APPLICANT: LI, Xiao-Jiang
; APPLICANT: Blackshaw, Seth
; TITLE OF INVENTION: AMIORIDE-SENSITIVE SODIUM CHANNEL AND
; TITLE OF INVENTION: METHOD OF IDENTIFYING SUBSTANCES WHICH STIMULATE OR BLOCK
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Alletrettl, LTD
; STREET: 1001 G Street, N.W., Eleventh Floor
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20001-4597
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/376,362A
; FILING DATE: 23-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan A., Sarah
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107, 48125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 508-9100
; TELEFAX: 202 508-9299
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 698 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; Molecule type: protein
; HYDROTHERMAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Rattus rattus
; US-08-376-362A-20

```

```

Query Match 3.6%; Score 90; DB 1; Length 698;
Best Local Similarity 18.9%; Pred. No. 2,3;
Matches 102; Conservative 47; Mismatches 147; Indels 244; Gaps 25;

QY 23 IFVFNSTKYMSALNISTSLH-----YMN----- 47
Db 83 LVOFCNNNTTHGAIKLVCSKNNRKAIFMAVIMLCTFGMMYWPALLFEELYSTPVLN 142
QY 48 ---SKSLFPRKPLISLKLPTETELRIKEIKLDQ----- 81
Db 143 INLNSDKLVFPAYVCTLNPRYTE--IKEELEDRITEQTLFDLYKNSVTPQAGAR 200
QY 82 -----QIPRPTHNTTSATSTATILNR----- 108
Db 201 RSSRDLGAFPHRLQRLTTPPPIS--GRTANSSSVRDNNPQVDRKMKIGPOLNQ 258
QY 109 -----DTYRGDQHLILEVROHLGRNQYGDELARMSPPALMAGASG----- 153
Db 259 NKSDFCYQYVSSG---VDAAVREWY--RFHYINIILSRISDTPALDEEALGNFPTCRF 311
QY 154 -----TSQV--HS-----ECGLIINTNAELCOYLNRPOES---FY- 242
Db 312 NQAPCNQANYSKTHHPMYGNCYTFNDKNSNLMMSHPGVNNG---LSLRLTEQNDPLP 368
QY 179 LL-----LHPSE-----GVSAIWSARNOGYDRIFTGOF-----VNG- 211
Db 369 LLSTVYGARVYVGGDEPAFMDGGFNLPRGVETSIKMKKALDSL--GGNYGDCITENG 426
QY 212 -----TSQV--HS-----ECGLIINTNAELCOYLNRPOES---FY- 242
Db 427 DVPYKVLVPSKYTQYVCHISCPQENMIKKGCAYIFRPRGVCFQYKROSSWGYCYK 486
QY 243 -----CYRPHMPCALTHM---YSKNNKVSYSKOKSLSFERSNV--GVEIM 285
Db 487 LOGAFSLDSJGCSKCRKPCSVYKLSAGYSWPSV---KSDWIMFSLTQNNYIN 542
QY 286 EKFNITSVSKCNT-LKSYDLHESGKLOHOLAVLDLRNINIQMOKYCPILGSMYTSVEM 344
Db 543 NKRN--GVAKLNTFFRELNYKTNSESPSVIMVSLNLSQMSLM-----FGSVLSYVEM 596

RESULT 14
US-08-530-950-13
; Sequence 13, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel
; APPLICANT: Gupta, Shaahl
; APPLICANT: Delibard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983

```

REFERENCE/DOCKET NUMBER: 07917/010001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 668 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 US-08-530-950-13

Query Match 3.6%; Score 89.5; DB 1; Length 668;
 Best Local Similarity 18.8%; Pred. No. 2.4; Mismatches 134; Indels 109; Gaps 14;
 Matches 72; Conservative 69;

QY 60 ISLPLETELRLKEIIEKLDQIIPRPFT--HVNTTTS-----AT 98
 DB 230 MSLKMPKT-----AQPOQFAPSPSKKHIEITLSNKSVEGKRNPGSLINGVST 281
 QY 99 HSTATILNPRDTCRGDQLHILLEVRDHLGRKQYGDFLRARSPPALMAGASGVTD 158
 DB 282 STSSSTEGPHDT-----VGTTPRTG---NSNNSNGSSGGGGLPANE 321
 QY 159 NNGTYLVSFTLFWEGVSLILLHPSEGVSAIWSARNQYDRIYFTGQFVNGTSQVSE 218
 DB 322 SKYDIKSGSLNFRAGKLSLSSGIDFSGSSRTILDELEFLDELGHGNGNYSKYLKP 381
 QY 219 CGLLINT-----NAELCOYLDNRDQ-----EGFYCVRPQIMPCA 252
 DB 382 TNVIMATKEVRELDEAKFQIIMLEVLHKCNSPYIVDFGAFTEGAVYMCMEYMOGG 441
 QY 253 ALTHMYSKKNKVSYSKOE-----KSLFRSNVGEIEMKFNITSVSKCNILK 300
 DB 442 SLDKIYDESSSEIGIDEPQLAFIANAVIHGLKELKEQHNITHROVKFTNLLCSANGTVK 501
 QY 301 SYDLHESGKLQHLAVDLDNRINIQWOKCYF-LIGSM-----TYSKEMEYLTIAIDRT 354
 DB 502 LCDFGVSGNLVASLA-----KTNIGQSYMAPERIKSLNPDRAITYV-QSDIWSLGL--- 552
 QY 355 GGEKNYIVISLQHFRRPPIIDVF 378
 DB 553 -----SILEMALGRY--PYPPETY 569

RESULT 15
 US-09-149-879-13
 Sequence 13, Application US/09149879
 Patent No. 6174676
 GENERAL INFORMATION:
 APPLICANT: Davis, Roger J.
 APPLICANT: Raingeaud, Joel
 APPLICANT: Gupta, Shashi
 APPLICANT: Dettlard, Benoit
 TITLE OF INVENTION: CYTOKINE, STRESS-, AND
 TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
 NUMBER OF INVENTION: KINASES
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/149, 879
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/530, 950
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Fasse, J. Peter
 REGISTRATION NUMBER: 32,983
 REFERENCE/DOCKET NUMBER: 07917/010001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 668 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 US-09-149-879-13

Query Match 3.6%; Score 89.5; DB 4; Length 668;
 Best Local Similarity 18.8%; Pred. No. 2.4; Mismatches 134; Indels 109; Gaps 14;
 Matches 72; Conservative 69;

QY 60 ISLPLETELRLKEIIEKLDQIIPRPFT--HVNTTTS-----AT 98
 DB 230 MSLKMPKT-----AQPOQFAPSPSKKHIEITLSNKSVEGKRNPGSLINGVST 281
 QY 99 HSTATILNPRDTCRGDQLHILLEVRDHLGRKQYGDFLRARSPPALMAGASGVTD 158
 DB 282 STSSSTEGPHDT-----VGTTPRTG---NSNNSNGSSGGGGLPANE 321
 QY 159 NNGTYLVSFTLFWEGVSLILLHPSEGVSAIWSARNQYDRIYFTGQFVNGTSQVSE 218
 DB 322 SKYDIKSGSLNFRAGKLSLSSGIDFSGSSRTILDELEFLDELGHGNGNYSKYLKP 381
 QY 219 CGLLINT-----NAELCOYLDNRDQ-----EGFYCVRPQIMPCA 252
 DB 382 TNVIMATKEVRELDEAKFQIIMLEVLHKCNSPYIVDFGAFTEGAVYMCMEYMOGG 441
 QY 253 ALTHMYSKKNKVSYSKOE-----KSLFRSNVGEIEMKFNITSVSKCNILK 300
 DB 442 SLDKIYDESSSEIGIDEPQLAFIANAVIHGLKELKEQHNITHROVKFTNLLCSANGTVK 501
 QY 301 SYDLHESGKLQHLAVDLDNRINIQWOKCYF-LIGSM-----TYSKEMEYLTIAIDRT 354
 DB 502 LCDFGVSGNLVASLA-----KTNIGQSYMAPERIKSLNPDRAITYV-QSDIWSLGL--- 552
 QY 355 GGEKNYIVISLQHFRRPPIIDVF 378
 DB 553 -----SILEMALGRY--PYPPETY 569

Search completed: October 12, 2001, 16:03:51
 Job time: 74 sec

Fri Oct 12 16:09:16 2001

us-09-729-454-1.rai

Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: October 12, 2001, 16:02:37 ; Search time 42.34 Seconds

(without alignments)
1464.291 Million cell updates/sec

Title: US-09-729-454-1

Sequence: 2485 1 MKISMINKSLALLFLIAS.....VHPPOHVGNOINILNYIC 475

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Search: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database: SPTREMBL_16:
1: sp-archaea:
2: sp-bacteria:
3: sp-fungi:
4: sp-human:
5: sp-invertebrate:
6: sp-mammal:
7: sp-mice:
8: sp-organelle:
9: sp-phage:
10: sp-plant:
11: sp-rodent:
12: sp-unclassified:
13: sp-vertebrate:
14: sp-virus:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	340.5	13.7	160	4	Q9NXP5	Q9NXP5 homo sapien
2	129.5	5.2	744	11	Q9ESN6	Q9ESN6 mus musculu
3	127.5	5.1	792	4	O60272	O60272 homo sapien
4	119.5	4.8	3603	2	P94459	P94459 bacillus su
5	112.5	4.5	4060	14	Q9IH28	Q9IH28 gill-assoct
6	109	4.4	688	5	Q9TY27	Q9TY27 caenorhabd
7	108.5	4.4	1387	5	O9G276	O9G276 plasmodium
8	106	4.3	2571	2	O87704	O87704 bacillus su
9	104.5	4.2	3262	4	O13788	O13788 homo sapien
10	104	4.2	728	4	O13779	O13779 homo sapien
11	102	4.1	800	5	O43988	O43988 dictyostel
12	99.5	4.0	602	5	O9VS73	O9VS73 drosophila
13	98.5	4.0	1507	6	O9S151	O9S151 canis famli
14	98.5	4.0	795	10	Q9STU3	Q9STU3 arabisopsi
15	98.5	4.0	1004	5	O96167	O96167 plasmodium
16	98	3.9	470	3	O12049	O12049 saccharomyc
17	98	3.9	657	2	O9PN6	O9PN6 campylobact
18	98	3.9	867	13	O9W65	O9W65 xenopus lae
19	98	3.9	3699	3	Q9HFB8	Q9HFB8 schizosacch

20	97.5	3.9	953	4	Q9Y239	Q9Y239 homo sapien
21	97	3.9	538	14	O83653	O83653 mumps virus
22	97	3.9	897	10	O41865	O41865 zea mays (m
23	97	3.9	1366	2	O54165	O54165 shigella fl
24	97	3.9	1802	5	O96170	O96170 plasmodium
25	96.5	3.9	1245	3	O03868	O03868 saccharomyc
26	96	3.9	538	14	Q9JAE8	Q9JAE8 mumps virus
27	96	3.9	952	5	Q9VX01	Q9VX01 drosophila
28	96	3.9	2692	5	O21547	O21547 caenorhabd
29	95.5	3.8	1285	14	Q9HP2	Q9HP2 yaba-like d
30	95.5	3.8	1685	10	O43248	O43248 zea mays (m
31	95	3.8	610	5	O27128	O27128 wechis cau
32	95	3.8	851	14	Q9V584	Q9V584 fowlpox vir
33	95	3.8	1021	5	O15733	O15733 dictyostel
34	94.5	3.8	740	10	Q9SVH0	Q9SVH0 arabidopsi
35	94.5	3.8	803	2	O82879	O82879 streptococ
36	94.5	3.8	2151	14	Q9DYJ9	Q9DYJ9 hantavirus
37	94	3.8	538	14	O9NAT0	O9NAT0 mumps virus
38	94	3.8	971	2	O51579	O51579 borrelia bu
39	94	3.8	1286	14	O9G903	O9G903 Shope fibro
40	94	3.8	3495	2	O30880	O30880 schizosacch
41	93.5	3.8	650	5	Q9U085	Q9U085 xenopus lae
42	93.5	3.8	684	4	Q9UKF6	Q9UKF6 homo sapien
43	93.5	3.8	1209	5	P91581	P91581 clona inte
44	93.5	3.8	1311	5	O9TYW6	O9TYW6 caenorhabd
45	93.5	3.8	2894	1	O58791	O58791 methanococ

ALIGNMENTS

RESULT	1	PRELIMINARY	PRT	160 AA.
Q9NXP5	Q9NXP5			
AC	Q9NXP5			
DT	01-OCT-2000 (TREMblrel. 15, Created)			
DT	01-OCT-2000 (TREMblrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMblrel. 15, Last annotation update)			
DE	CDNA FLJ20127 FIS, CLONE COL06176.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
RN	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLON;			
RA	Kawabata A., Hiki J. T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,			
RA	Okutani R., Oca T., Suzuki Y., Obayashi M., Mishi T., Shibahara T.,			
RA	Tanaka T., Nakamura Y., Isogai T., Sugano S.;			
RT	"NEDO human cDNA sequencing project."			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AK000134; BAA0965.1; "			
SO	SEQUENCE 160 AA; 18466 MW; 113114CA43912208 CRC64;			

Query Match	13.7%	Score 340.5;	DB 4;	Length 160;
Best Local Similarity	51.0%;	Pred. No. 4.1e-20;		
Matches	73;	Conservative	0;	Mismatches 1;
			Indels	69;
			Gaps	1;
QY	285 MEKFNITSVSKC-----			296
DB	1 MEKFNITSVSKC-----			60
QY	297 -----			335
DB	61 LITLMDSTIRQMEYTKASINTLSVDHESGKLQHLQAVLDLRNINIQOKYCPFLIG 120			
QY	336 SMTYSVKEKEYLTRAIDRTGEX 358			
DB	121 SMTYSVKEKEYLTRAIDRTGEX 143			
RESULT	2			

Q9ESN6
ID Q9ESN6 PRELIMINARY; PRT; 744 AA.
AC Q9ESN6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NEURAL ACTIVITY-RELATED RING FINGER PROTEIN.
GN NARF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR MICE;
RA Ohkawa N.;
RT *cDNA sequence encoding Neural Activity-Related Ring Finger Protein (NARF).
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB043550; BAB17634.1;
SO SEQUENCE 744 AA; 81445 MW; 80336DA2EPESEFD CRC64;

Query Match 5.2%; Score 129.5; DB 11; Length 744;
Best Local Similarity 22.1%; Pred. No. 0.051;
Matches 89; Conservative 43; Mismatches 133; Indels 137; Gaps 18;

QY 66 TETELRI--KEIIEKL-----DQIIPRP-----FTHYNTTSA 97
DB 266 TETELVLYVKKMSKLNELADDFLHPRENDQDLYVETELKKSHINLTITLITNAVA 325
QY 98 THSTATILNPDIYCGDQDLHILEVRDHLGRKQYGGFLFARMSPALMAGASGVTD 157
DB 326 SETVATGEGELQGITI-GQPMSTVITTKDKDELCKGNAYLAEISTPD-GVAADGEILD 383
QY 158 FNNGYLVSFTLFWEGVSLSLT-----IHP-SEGVALMSARNG 197
DB 384 NKGTYEFLTYVKEGDFLTLRLDQHIRSGPKLKYIRSDVSPTEGYKKRVKSPGS 443
QY 198 GY-----DVIYF-----TGCFVNGTSQVHSECG--LIL 223
DB 444 GHVKQKAVKRPASMYSTGRKKNPEIDDLIFRVGTGRNKGFEITLQGVAASTSGKILIA 503
QY 224 NTNAELCYLDNRDQ-EGFYCVRPQ-----HMPCAALTHMS-----KNKKVSYLSKQ 270
DB 504 DSNNGCVQIFSNQGFKSRFGIRGSRPQQLQRTGYAVHPSDIIADYONKMWSTFSSD 563
QY 271 EKSLFERNNGVGEIMEKFNITISVSKNTLASVDLHESGKLQHLAVDLDRN--INTQWK 328
DB 564 GRF---KTKIG-----SGKLMPKGVSVDRNGHIIVYDNK 595
QY 329 YCYPPLIGSMYSVYKEMEYLTRAIDRTGGEKNVTYVLSLQHF 370
DB 596 ACCVFIFQPNKGI-----VTRFGSRNGNDQGA-----GPHF 627

RESULT 3
ID 060272 PRELIMINARY; PRT; 792 AA.
AC 060272;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE KIAA0517 PROTEIN (FRAGMENT).
GN KIAA0517.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98290545; PubMed=9628581;

RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nishita N.,
RA Ohara O.;
RT *Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.*;
RL DNA Res. 5:31-39(1998).
CC -!- SIMILARITY: CONTAINS A RING (CHRC4-CLASS) ZINC FINGER.
DR EMBL; AB011089; BAA25443.1;
DR HSSP; P28990; ICHC;
DR InterPro; IPR000315;
DR InterPro; IPR001258;
DR InterPro; IPR001298;
DR InterPro; IPR001841;
DR InterPro; IPR002991;
DR InterPro; IPR003649;
DR Pfam; PF00097; ZF-C3HC4; 1.
DR Pfam; PF00630; Filamin; 1.
DR Pfam; PF00643; Zf-B_box; 1.
DR Pfam; PF01436; NHL; 6.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR SMART; SM00502; BRC; 1.
KW Zinc-finger.
FT NON-TER
SO SPOUDENCE 792 AA; 86659 MW; DA01F403BE580908 CRC64;

Query Match 5.1%; Score 127.5; DB 4; Length 792;
Best Local Similarity 22.1%; Pred. No. 0.081;
Matches 89; Conservative 43; Mismatches 133; Indels 137; Gaps 18;

QY 66 TETELRI--KEIIEKL-----DQIIPRP-----FTHYNTTSA 97
DB 314 TETELVLYKKMSKLNELADDFLHPRENDQDLYVETELKKSHINLTITLITNAVA 373
QY 98 THSTATILNPDIYCGDQDLHILEVRDHLGRKQYGGFLFARMSPALMAGASGVTD 157
DB 374 SETVATGEGELQGITI-GQPMSTVITTKDKDELCKGNAYLAEISTPD-GVAADGEILD 431
QY 158 FNNGYLVSFTLFWEGVSLSLT-----IHP-SEGVALMSARNG 197
DB 432 NKGTYEFLTYVKEGDFLTLRLDQHIRSGPKLKYIRSDVSPTEGYKKRVKSPGS 491
QY 198 GY-----DVIYF-----TGCFVNGTSQVHSECG--LIL 223
DB 492 GHVKQKAVKRPASMYSTGRKKNPEIDDLIFRVGTGRNKGFEITLQGVAASTSGKILIA 561
QY 224 NTNAELCYLDNRDQ-EGFYCVRPQ-----HMPCAALTHMS-----KNKKVSYLSKQ 270
DB 552 DSNNGCVQIFSNQGFKSRFGIRGSRPQQLQRTGYAVHPSDIIADYONKMWSTFSSD 611
QY 271 EKSLFERNNGVGEIMEKFNITISVSKNTLASVDLHESGKLQHLAVDLDRN--INTQWK 328
DB 612 GRF---KTKIG-----SGKLMPKGVSVDRNGHIIVYDNK 644
QY 329 YCYPPLIGSMYSVYKEMEYLTRAIDRTGGEKNVTYVLSLQHF 370
DB 644 ACCVFIFQPNKGI-----VTRFGSRNGNDQGA-----GPHF 675

RESULT 4
ID P94459 PRELIMINARY; PRT; 3603 AA.
AC P94459;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PEPTIDE SYNTHETASE ORF4.
GN pps4 OR ppsd.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
OX NCBI_Taxid=1423;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-9522736Z; PubMed-7711903;
 RA Tomoni A., Franchi E., Magistrelli C., Colombo E., Cosmina P.,
 RA Grandi G.,
 RT "A putative new peptide synthase operon in *Bacillus subtilis*: partial
 RT characterization."
 RI Microbiology 141:645-648(1995).
 RN 12.
 RC SEQUENCE FROM N.A.
 RP STRAIN-168;
 RA Grandi G.,
 RN Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 RC SEQUENCE FROM N.A.
 RP STRAIN-168;
 RA De Ferra F., Tognoni A.,
 RN Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-96044033; PubMed-9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertier M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Boriss R., Bouvier L., Baus A., Braun M., Bruppel S.C., Bron S.,
 RA Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codan J.J., Conerton J.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Enliat K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C.D., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Kleier-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauder J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogilwa A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Pujic P., Puttelie B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemura K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
 RA Viari A., Wanduit R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the gram-positive bacterium *Bacillus*
 RT *subtilis*."
 RI Nature 390:249-256(1997).
 RN 15).
 RC SEQUENCE FROM N.A.
 RP STRAIN-168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,
 RN Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE OF 3071-3603 FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-98048468; PubMed-9387222;
 RA Tosato V., Albertini A.M., Zotti M., Sonda S., Brunsch C.V.,
 RT "Sequence completion, identification and definition of the fengycin
 RT operon in *Bacillus subtilis* 168."
 RI Microbiology 143:3443-3450(1997).
 RN 16.
 RC EMBL: 234883; CA84363.1;
 DR EMBL: 299113; CA13714.1;
 DR EMBL: Y13917; CA74212.1;
 DR HSSB: P14687; LAMU;
 DR InterPro: IPR000255;
 DR InterPro: IPR000873;
 DR InterPro: IPR001242;

DR Pfam: PF00501; AMP-binding; 3.
 DR Pfam: PF00550; pp-binding; 3.
 DR Pfam: PF00668; Condensation; 4.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00075; ACP_DOMAIN; 3.
 DR PROSITE: PS00455; AMP_BINDING; 3.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
 KW Phosphopantetheine.
 SQ SEQUENCE 3603 AA; 406807 MW; 9EC0486E9229D8 CRC64;
 Query Match 4.8%; Score 119.5; DB 2; Length 3603;
 Best Local Similarity 20.7%; Pred. No. 3.2;
 Matches 98; Conservative 73; Mismatches 149; Indels 153; Gaps 25;
 QY 23 IFVYVNSTVY-----WSALNLSILHYMNSKSLFPPKPLSLKPLTELERIKE 74
 DB 1426 LFCQFEVSTALFEKETIQHWASLFTLVH-----TASPETELDIPILKEER--D 1477
 QY 75 IIEK-----LDQIIRPP-----FTHWTTTSAHSTAPIL 105
 DB 1478 FIESCHLFETGYSMNQTLHTALDAQEKTPQAAVIEDGVMYTELNEQATRIEMELI 1537
 QY 106 NERDTYGRGDHILLEVADHIGRRKQYGGDFELARMSSPALMAGSGVTLFNNGYLY 165
 DB 1538 G-----RGVPEYVAI---IGKR-----SPMLIGITGLK--AGAVLP 1573
 QY 166 SFLLFMGCVSYLLIHPSEGVSALMSARNOGYDRVFTGQYNGTSQVHSEGLIANT 225
 DB 1574 IDPPYPERI--SFLEDSGNILLQSA---GLHVPEFGEIV-----YLMOT 1617
 QY 226 NAEICQYLDNRQDESPYVPRQHPNCAALTHMYSKKKVSYLSKQESLFERSNVGEIM 285
 DB 1618 NSGLAHLSPNVD---VLQSLAYIYTSSTQMPK-----GVIE 1656
 QY 286 EKFTIVSKCNTLKS-VDLHESGKLQHLQAVLDNRNINOMOKYCPILGMYTSV--- 341
 DB 1657 HR--SAVNINLSISQRYQLKHSMDIMKITSFPASt--WELFWPPAGASVYLIPQG 1710
 QY 342 --KMETLRADRFEGGKNTVYISLQGFPPFIDVIRRALNKAIOHLNRSPT 399
 DB 1711 GKKPEVKAIEE--OKTAM-----HFVPSML-----HALEQIKYRS--- 1748
 QY 400 WYIKTENIREMYNDAREFSDFGYIOYLIDIPDOLSYSIIDAMDIANG 452
 DB 1749 -VPKTNRLKRVFSGSGGLG-----THLYSR--FYELLRV-----SITNSIG 1788
 RESULT 5
 QYIH28 PRELIMINARY; PRT; 4060 AA.
 AC QYIH28;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE ORF1A POLYPEPTIDE.
 OS gill-associated virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales.
 RN NCBI_TaxID-96491;
 RX SEQUENCE FROM N.A.
 RC MEDLINE-20273993; PubMed-10811931;
 RA Cowley J.A., Dimmock C.M., Spann K.M., Walker P.J.,
 RT "Gill-associated virus of penaeus monodon prawns: an invertebrate
 RT virus with ORF1A and ORF1b genes related to arteri- and
 RT coronaviruses."
 RI J. Gen. Virol. 81:1473-1484(2000).
 RN EMBL: AF227196; AAF82689.1;
 DR EMBL: AF227196; AAF82689.1;
 DR InterPro: IPR000923;
 DR PROSITE: PS00196; COPPER_BLUE; UNKNOWN_1.
 KW Polyprotein.
 SQ SEQUENCE 4060 AA; 460302 MW; ADC47C9A9A817C2 CRC64;

Query Match 4.5%; Score 112.5; DB 14; Length 4060;
 Best Local Similarity 19.6%; Pred. No. 14;
 Matches 103; Conservative 85; Mismatches 213; Indels 125; Gaps 26;

QY 22 IIFVFNSTKYWSALNSISLHWNNSTKSLPPTPLISIKPLETELRIKEIKID-80
 DB 1614 IETDLDKST--FTALYHFIKHY--NFKYTPQPPVCKTKLMMPTPTQYISDCDTQEL 1668
 QY 81 -----QQIPRPFTHVNTT-----SATHSTATILNPR-----108
 DB 1669 GPKHINFEQVYDTRRNITLHNVPTQSDIVFTTTTEYDSNHOAVTDRFTIN 1728
 QY 109 -DYCRGDOLHILEVRDHLGR-KYGGDFLRAMSSPALMAGASGVTDGNNGTLYVS 166
 DB 1729 STTTLPSDLTLRAQNRNKKOSRMHNYGHRFYNAELLKPGTNG--HMIPSLGGPAAVA 1786
 QY 167 FTLEMGVSLSLILHPSGVSALWSANQGY--DRVFTGQ--FVNGTSQVH--216
 DB 1787 NAFLQTIQANLD-DTYENQAINAVYNNKRIYVYNDIIGDASDWCYINSPSPPTL 1845
 QY 217 -----SECGILNTNALCOYLDND-----QEGFTCVRPOHMPALITHWYSKNKVVS 265
 DB 1846 ALPRYSSEAEFTLTHGNI--YVTHDHYHTEEFELIRSLALPGVYSAISLNGE--1901
 QY 266 YLSKQKSLFERSNVGEVIMEKENTISVS---KCNTELSDVLESGLQHLAVDLDNR 321
 DB 1902 --TVRRMNPSDQCGKTLHHPINTHTHWTMLATNLKTA-ITAAKLRHPALATL--G 1956
 QY 322 INIOMOKYCYPL-----IGSNITSYKEMEYLRADRTGSEKNTVYVSLQGHFRPPID 376
 DB 1957 LSPDFLTAHPHRIHTLGS--YTHHNLAYLLPALRCA---SGVAVFGVN--PPLN 2008
 QY 377 VFIRALNVHKAIOHLRSPDTMVIKTE-----NIREM--YNDARESDFGY 424
 DB 2009 CTSDKGYINGSLPSEHNTQNFATINPEPTTYSGLVRMTCOLYSIRLWADNGY 2068
 QY 425 IQYIIKIDIFODLSIIDAMDTIYKGNVH-----PPQH 461
 DB 2069 -----STSTDTADITFRF--NNSVPTRCYSYKIPPTH 2100

RESULT 6
 Q9Y27 PRELIMINARY; PRT: 688 AA.
 AC Q9Y27;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE F58E2.3 PROTEIN.
 GN F58E2.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copesey J., Cooper J., Coulson A.,
 RA Craixton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latteille P.,
 RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Koopra A., Saunders D., Showkeen R.,
 RA Shalson N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstein L., Wilkinson-Sprat J., Wohlman P.,
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL elegans.*
 RN Nature 368:32-38(1994).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Goela D., Delehaunty A.,
 RT "The sequence of C. elegans cosmid F58E2."
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.,
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF100659; AAC68966.1;
 DR Interpro: IPR001810;
 DR Interpro: IPR002900;
 DR Pfam: PF00646; F-box; 1.
 DR Pfam: PF01827; DUF38; 2.
 SQ SEQUENCE 688 AA; 79110 MW; 6EA4BCD3244CA1A CR364;

Query Match 4.4%; Score 109; DB 5; Length 688;
 Best Local Similarity 19.3%; Pred. No. 2.1;
 Matches 112; Conservative 76; Mismatches 200; Indels 192; Gaps 27;

QY 1 MKTSMINYSLLALLFLTASWILFTVFNSTKYWSALNSISLHW-----NNSIKSLF 64
 DB 93 LEISQDIEEL-----DKRNASASSQSIOTYDFIOTPLNHRITGK 144
 QY 55 PKTPILSL-----KPLETELRIKEIKLQUTPR 66
 DB 134 PK--LISLFGTYQALSFRIHNSNHEIINLNRKREDELEIHFIEHIVM--KMKRHA 169
 QY 87 PFTHVNTTSATHSTATI--LNPDTYCGGQDHLILEVRDHLGRKLYGGDFLRAMSS 144
 DB 190 KYHLEECTSTENVLHFGANRKYDVISIDLDLMELSTHAIKIPNISF-----242
 QY 145 PALMAGASKVTD-----FNNGTIVSFTLW--EQVSLSLILHISEGV--SALWS 194
 DB 243 ---NNAADSIIDSLGLFGSPKSLFSEWMSIDKIECSFSEDFKFSIYCPKACE 298
 QY 194 ARNGYDRVFTFGFVNGTSQVHSEC-----GLINT-----NAELQYLDNRQGB 240
 DB 299 LRN-----YQSSKELECTSLCTSELSTIKVLAHKKLKNLISLGHIDICL--KRSK 250
 QY 241 FYCVRQHPFCAALITHWYSKNKVSYLSKQKSLFERSNVGEVIMEKENTISVSK-----296
 DB 351 V-----QKCVDI-----KDEEIIAIEINJENNVSAJ---FNAHKKAYEP 394
 QY 297 -----NTLKSVDLESGLQHLAVDLDNRNINQW-----KTYPLIUSMT 348
 DB 395 IDLTTSYQDHSKYIQKVEKHNGLVERELC-----SVNKKGLQPSLKAHILKINM 444
 QY 339 VSYKEME-----YLRADRTGGEKNTVYI-----STQGHFRPPIDIVT 479
 DB 450 AOKSEIQVRLDFSLTKANSRQSKKEHVLLFRLVAQOLDVYSTIQLQINQTPYKKL 509
 QY 380 RRALNVHKAIOHLRSPDTMVIKTEINIREMYNDARESDFGYIYLLIKLIFGLAS 499
 DB 510 IQYLNQETLIQLSLPSS--LKTIEIRLKN-----IISKRLKLNKA 553
 QY 440 SIIDAM-----DITATGNTNVPQ-----HVCNQINIL 470
 DB 554 AQDDQKKAEDILEVLT--APGQIQYELHFVNTTEINLM 590

RESULT 7
 Q9G276 PRELIMINARY; PRT: 1387 AA.
 AC Q9G276;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE RETICULOCYTE-BINDING PROTEIN 2 HOMOLOG A (FRAGMENT).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20402589; PubMed=10920203;
 RA Hayer J.C., Galinski W.R., Ingravallo P., Barnwell J.W.;
 RT "Two plasmodium falciparum genes express mezozoite proteins that are
 related to Plasmodium vivax and plasmodium yoelli adhesive proteins
 involved in host cell selection and invasion."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:9648-9653(2000).
 DR EMBL: AF196347; AAF98066.1;
 FT NON_TER 1
 FT SEQUENCE 1387 AA; 164871 MW; 5660154744FC285 CRC64;

Query Match 4.4%; Score 108.5; DB 5; Length 1387;
 Best Local Similarity 20.2%; Pred. No. 6.4;
 Matches 80; Conservative 70; Mismatches 137; Indels 109; Gaps 19;

QY 140 ARMSPALMAGASCXYDEN--NGTIVLSTFLPREGVSLSLIHSE--GVASALMSARN 196
 DB 250 ANFGSGLIYMSDITETTPENPLENDLLNQLFPERKEHTSTLENSDLELDHLSNSD 349
 QY 197 GGTDRVLTGQFVNGTSGVHSEGLLNTNAELQYLNRDQEGFVCPQHPQCAL-- 254
 DB 350 ESIUNL---KYVNDLIEIHLT-----YSTULKYLNDICKLKGDG--NDLVKCKELRE 397
 QY 255 --THMYSKKNKSVLSKQESL-----FERSNVGVEIKENFNTIS 292
 DB 398 LSTALYDLKIOITSVINRENDISNNLIDIAIQVNEFKYKEIFPDVNEEYKTLD 457
 QY 243 -----VSKCNTLSVDLHESGKLOHQLAV---DLDRNINIQOKYCPYLGSMYSVK 342
 DB 458 DTKNAVIVKKAELKNVDIN---KTKREDIDYFNLDL-----LEKSLTSSN 502
 QY 343 EMEYLT-----RAIDRTGSEKNIVIVSLGQFRPPIDVFIRALVWKAI 389
 DB 503 EMEKTIYVNSYNSFSIDINKNIDKEKTLIPN-----LDELINCHNIDISL 552
 QY 390 QHLLRSPTWVITKTEINREMYNDARESDPHGTYQ--YLIKIDFQDLSVSIIDAMD 447
 DB 554 YNFILNIOIKIKGNDIKIKRQENDTNICFE---YIDNNNFILKS-----DISIFNKYDD 604
 QY 448 TIA---YGTNNVHP-----POHVGNOINILN 472
 DB 605 HIKVDNYISNNIDVNNKNSLSEHVI--NATNILEN 639

RESULT 8
 ID 087704 PRELIMINARY; PRT: 2571 AA.
 AC 087704;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE FENGYCIN SYNTHETASE.
 GN FEND.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=F29-3;
 RA Huang S., Chen C., Lui S.;
 RT "Analysis of fengycin synthetase gene fend.*;
 Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A1011849; CAA09819.1;
 DR HSSP: P14687; JAM0;
 DR LOCUSPRO: IPRO00255;
 DR InterPro: IPRO00873;
 DR InterPro: IPRO01242;
 Pfam: PF00501; AMP-binding; 2.

DR Pfam: PF00550; pp-binding; 2.
 DR Pfam: PF00668; Condensation; 3.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00075; ACP-DOMAIN; 2.
 DR PROSITE: PS00455; AMP-BINDING; 2.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
 KW Phosphopantetheine.
 SQ SEQUENCE 2571 AA; 290799 MW; 8309833ADA88 CRC64;

Query Match 4.3%; Score 106; DB 2; Length 2571;
 Best Local Similarity 19.3%; Pred. No. 25;
 Matches 95; Conservative 71; Mismatches 172; Indels 154; Gaps 25;

QY 109 DTTCGQDQHLIL-LEVRD-----HLGR-----KQYGDPLFARMSPAL--- 147
 DB 1199 DLXV-GKELPALMLHYKDFAVWQOEKPEETHYSKQAVWLKQLG-----SLFVLDLP 1250
 QY 148 MAGASGKVTDFNNGTYLVSF-----TLFREGVSLSLIHSEGVSLMSARNQG 198
 DB 1251 LDKTRPLLPFSGTTEVNIIDKRTAEELHRLMEFTGTMYMLL---AVYSIMLSKLSG 1306
 QY 199 YDRVLTGQFVNGTSGVHSE--CGLLNTNAELQY-----YLDNRDQGFYCPQ 247
 DB 1307 QEDIV-VGSPAAGRPHADLERVIGMFNTLAMKQCPBGRKTFSSYQELRELALAYEHQ 1365
 QY 248 HMPQALTHMYSKKNKV-----SYLSKQESLFEERNVGVLEMEKNTISVSKNTLK 300
 DB 1366 DYFPEELVNNLETKREVRNPLFPAALVYQNSEDQFVPELSTSSVTSVSHVNSFDTLT 1425
 QY 301 SVDLHESG---KLOHQLAVDDLDRNINIQOKYCPYLGSMYSV---KEMEYLAIDR 353
 DB 1426 HAEBHNSGRCREYSTALFEETI--ARASHIELVYGTISDIHKKLSEMOILS----- 1479
 QY 354 TGSEKNIVIVSLGQFRPPIDVFIRALN-----YHKAI--QHLL----- 394
 DB 1480 --AFARELLETWQO--YADYPRDESIVRLFEGQAHEHHTAVVCHILTYRLDNKAE 1536
 QY 395 -----RSPDTWV---ILKT-----ENIREMYNDA 415
 DB 1537 RTATMLMKQVSGDIIQLMFDRSPDVTGCVSLTKMGAVLPIDEPYPOEISPMND 1596
 QY 416 -----ERSEDFHYIYLIKIDFQDLSVSIIDAMDIT-----IAYGTNNVHP 458
 DB 1597 GARILLTEGHNNKPADYHQIYLV--NDAENELISPLKQAEITLADQPAYIYISGTTQ 1654
 QY 459 POHVGNOINIL 470
 DB 1655 PKGVVEHRNVI 1666

RESULT 9
 ID 013788 PRELIMINARY; PRT: 3262 AA.
 AC 013788;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE APOLOPROTEIN B-100 (FRAGMENT).
 GN APOB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87191999; PubMed=2883086;
 RA Carlsson P., Darinfor C., Olofsson S.O., Bjurzell G.;
 RT "Analysis of the human apolipoprotein B gene: complete structure of
 the B-74 region."
 RL Gene 49:29-51(1986).
 DR EMBL: M15421; AAAS1758.1;
 FT NON_TER 1

SEQUENCE 3262 AA; 370140 MM; 56603BC0618DD40D CRC64;

Query Match 4.2%; Score 104.5; DB 4; Length 3262;
Best Local Similarity 20.5%; Pred. No. 46;
Matches 89; Conservative 72; Mismatches 169; Indels 105; Gaps 22;

```

OY 56 KTLPLSLKPLTEPE-LRIKEIIEKIDDOIPRPPTHV-----NTTSATHS-TATILNPRD 109
DB 744 KVPLLSEPINIDALEMRDAVEK-----POEFTIVAFVKYKQNOVHSINLPFEFTIQ 797
OY 110 TYCRGDQHLILEVRDHLGRKQYGD-FLPARSSPALMAGSKGYDPFNGIYLVST 168
DB 798 EYERNRQTIIVLENVQRNLKHINIDQFVK-----YRAALKLQCAN-DYLSNEN 849
OY 169 LMEGOVSLSLILHPSEGSALMS-----ANAGYDRAVF 204
DB 850 --WENQVS-----HAKKLTALTAKRYITENDIQALDANKINENKLSQLOTY--MIQ 899
OY 205 TCGFVNGTSQVHSECGIILNTNAELCOYLDNRDQEGFYCVPRQHPGCAALTHMYSK---- 260
DB 900 FQOYIKSDYDLHDKLAIANIIDEIIEKLSLDEH--YHIRVNLVKTTHDLHLEFINDF 957
OY 261 NKKVSYLSKQKSLFERSNVGEIEMKENTISVKCNTLSYDL-HSGKL-QHOLAVDL 318
DB 958 NKSGSTASWQNVDTKYQIRIQIEKLOQK---RHQINIDIOHLAKLQKHEAIDV 1013
OY 319 DNINIQOKCYPLIGSMTYSVKEMEYLRADRTGEGKNTVVISLQHFPRPTI--- 375
DB 1014 -----RVLLDQLG-----TISFERI---NDVL-----EHVNHVYINLI 1044
OY 376 -DVFIIRALNVHKAIOHLLNSPOT-----MYIKTENIREMYDAREFSDHFQYLYLI 429
DB 1045 GDFEVAEKINAFKAKYHELIEREYDQIOVLMDKLVELAQYKELTKTQKLSNVLOQVK 1104
OY 430 IKDIFQDLVSITDA 444
DB 1105 IKDYFEKLVGFIDA 1119

```

RESULT 10
ID 013779 PRELIMINARY: PRT: 728 AA.
AC 013779;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE APOLIPROTEIN B-48 (FRAGMENT).
GN APOB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_Taxid=9606;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=88050832; PubMed=3676265;
RA Hardman D.A., Procter A.A., Chen G.C., Schilling J.W., Sato K.Y.,
RA Lau K., Yamana K.M., Mikita T., Miller J., Crisp T.M., McEnroe G.,
RA Scarborough R.M., Kane J.P.;
RT "Structural comparison of human apolipoproteins B-48 and B-100.";
RL Biochemistry 26:5478-5486(1987).
DR EMBL: M17367; AAA51741.1;-
FT NON_TER 1
FT NON_TER 728
SQ SEQUENCE 728 AA; 83217 MM; CF27745EA02B699 CRC64;

Query Match 4.2%; Score 104; DB 4; Length 728;
Best Local Similarity 20.5%; Pred. No. 5.8;
Matches 89; Conservative 72; Mismatches 167; Indels 106; Gaps 23;

```

OY 56 KTLPLSLKPLTEPE-LRIKEIIEKIDDOIPRPPTHV-----NTTSATHS-TATILNPRD 109
DB 744 KVPLLSEPINIDALEMRDAVEK-----POEFTIVAFVKYKQNOVHSINLPFEFTIQ 797

```

```

DB 355 KVPLLSEPINIDALEMRDAVEK-----POEFTIVAFVKYKQNOVHSINLPFEFTIQ 408
OY 110 TYCRGDQHLILEVRDHLGRKQYGD-FLPARSSPALMAGSKGYDPFNGIYLVST 168
DB 409 EYERNRQTIIVLENVQRNLKHINIDQFVK-----YRAALKLQCAN-DYLSNEN 460
OY 169 LMEGOVSLSLILHPSEGSALMS-----ANAGYDRAVF 204
DB 461 --WENQVS-----HAKKLTALTAKRYITENDIQALDANKINENKLSQLOTY--MIQ 510
OY 205 TCGFVNGTSQVHSECGIILNTNAELCOYLDNRDQEGFYCVPRQHPGCAALTHMYSK---- 260
DB 511 FQOYIKSDYDLHDKLAIANIIDEIIEKLSLDEH--YHIRVNLVKTTHDLHLEFINDF 568
OY 261 NKKVSYLSKQKSLFERSNVGEIEMKENTISVKCNTLSYDL-HSGKL-QHOLAVDL 318
DB 569 NKSGSTASWQNVDTKYQIRIQIEKLOQK---RHQINIDIOHLAKLQKHEAIDV 624
OY 319 DNINIQOKCYPLIGSMTYSVKEMEYLRADRTGEGKNTVVISLQHFPRPTI--- 375
DB 625 -----RVLLDQLG-----TISFERI---NDVL-----EHVNHVYINPY 655
OY 376 -DVFIIRALNVHKAIOHLLNSPOT-----MYIKTENIREMYDAREFSDHFQYLYLI 429
DB 656 GDFEVAEKINAFKAKYHELIEREYDQIOVLMDKLVELAQYKELTKTQKLSNVLOQVK 715
OY 430 IKDIFQDLVSITDA 443
DB 716 IKDYFEKLVGFIDA 728

```

RESULT 11
ID 043988 PRELIMINARY: PRT: 800 AA.
AC 043988;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HOMEBOX-CONTAINING PROTEIN WARIAL (FRAGMENT).
GN WARIAL.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
NCBI_Taxid=44689;
RN [1]
RX SEQUENCE FROM N.A.
RP STRAIN-KAX3;
RA Han Z., Firtel R.A.;
RL Development 0:0-0(1998).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER HOMEBOX DOMAINS.
DR EMBL: AF036170; AAB92245.1;-
DR HSSP: P42711; 1817;
DR InterPro: IPR001356;-
DR InterPro: IPR002110;-
DR Pfam: PF00023; ank; 9;
DR Pfam: PF00046; homeobox; 1;
DR PROSITE: PS50088; ANK_REPEAT; 8;
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1;
DR PROSITE: PS50027; HOMEBOX_1; 1;
DR PROSITE: PS50071; HOMEBOX_2; 1;
DR SMART: SM00389; HOX; 1;
KW DNA-binding; Homeobox; Nuclear protein.
FT NON_TER 800
SQ SEQUENCE 800 AA; 88723 MM; 7FACC2C09884DB4F CRC64;

Query Match 4.2%; Score 104; DB 5; Length 800;
Best Local Similarity 18.4%; Pred. No. 6.7;
Matches 83; Conservative 68; Mismatches 136; Indels 164; Gaps 19;

```

OY 104 IINPDP-----TYCRDQHLILEVRDHLGRKQYGD-FLPARSSPALM-AGASGXVD 157
DB 343 LHPNINIDFLIRINPDPSIDVARNDSKGLSLFLTAFLAGEYQVRLIESANINIKD 402

```

```

OY 158 FMNGTYLSTFLMEGVSLSLLIHNS-----EGVSALMSARNOG-----YD 200
1b 403 NOGDTYLAASVL-GNOPVELLEHRAEDNLVNDGVSPLFSACKGHQIXSSLLDD 461
OY 201 RVIFTGQFVNGTSQVH-----SECGLLNTNALCYLDNRDQGFYCAPQHPMCA 252
Db 462 REVSVKTKINGETPLHIAKCKFEICKLLETEAK-ASVIDSNRR-----TPLHAC- 513
OY 253 ALTHYKNNKVSYSLSKOKSLFERNVGEIEMKNTISVSKCNTKLSYDLHESKLOH 312
Db 514 -----IMGYSTAKLLJCN-----GADNN----- 532
OY 313 QLAVDLRNINIQOKYCYPLIGSKTYSKEMEYLTRAIDRTGKEKNTYVLSGQHPRP 372
Db 533 --AIDIDGHTPLH-----TSSLGQYLTITRLLENGADPN-----IDDSGCTYP 574
OY 373 FPIDVFRRLALNVHKAIOHLLRSP-----DTMVITIKENIREMYADARF----- 418
Db 575 -----ITHAVRESRIETVYKFLIKFNKSLIKTKNGQNLHLISVQFASLMMGQM 622
OY 419 -----SDFHUY-----IOYLIIRKD-----IFODLSYSI 441
Db 623 IFESKGEIADSDSGYTPYLAAGAKTNEVYKLLSKGSKKIRPLEKIQENDKEI 682
OY 442 IDAMDITAYGTNNVHPQHVAGNOINILN 472
Db 683 IOMLESTVTKSSNNNSNSNT--NNNNNNN 711

RESULT 12
OYVS73 PRELIMINARY: PRT: 602 AA.
AC OYVS73;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE CG8546 PROTEIN.
GN CG8546.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Morlan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer J.R., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Ayll J.F., Abdayani A., Au H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Daveport L.B., Davies P.,
RA de Pablos H., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunnov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
RA Fouts C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Grudok A., Gong F., Gotrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel R.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko J., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei J., McIntosh T.C., McLeod M.P., McPherson D.,
RA Morkulov G., Mishina N.V., Modarity C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

```

```

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pachet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D., Schaefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT The genome sequence of Drosophila melanogaster.*;
RL Science 287:2185-2195(2000).
DR EMBL: AE003559; AAF50551;
DR Flybase: FBgn0035785; CG8346.
DR InterPro: IPR001873; -.
DR Pfam: PF00858; ASC: 2.
DR PRINTS: PR01078; AMINACHANNEL.
SQ SEQUENCE 602 AA; 68399 MW; 7CDA34CC9D78A22A CRC64;

Query Match 4.1%; Score 102; DB 5; Length 602;
Best Local Similarity 20.9%; Pred. No. 6.4;
Matches 86; Conservative 61; Mismatches 146; Indels 118; Gaps 23;

OY 9 KSLALLFIASWI-----IFTVFQNSTKWSALNLSISLHYNN-----STKS 52
Db 122 RPLLERLF-----WISVLVSFTCVKLTILNMDK-----MNNPIYVFAKSTPV 168
OY 53 LFEPKPLISKLTELEIRKEIEIKEDQIIPPRF-THVTTTSATH-----STAT 104
Db 169 WQIFPAVATYCPETKTR---REIFNFDSTHOVROFQSNVGVYDSDRGHEIIDALTE 225
OY 105 LNPR--DTYCRGQQLHILEVRDHLGRKQYGDGLARRMSSPAMAGAGKVTDFNNGT 162
Db 226 VSPHFDDTYLNCMKWNSPVKCSDFHFHFTVDEGVCFSPNSLPAPFIRAEIGIIPF---- 281
OY 163 YLVSFTLF-MEGVSLSTLLIHPSSEGVSLMSARNOG-DEVIYFTQG-----FYNGS- 213
Db 282 -----IFRENRLSMDW---NVEDGISA--SADISPPNNVLPQARRAGLYLFWGAEI 330
OY 214 QVHSEC-----GLILNTNALCYLDNR-----DOGFYCVRPQHPMCAALTHMYSKN 261
Db 331 DFDDMCRGPQVGGKILLTPDGVAVOVSQVFRIPPDQVLSIRPKIITTSGLKHEPN 390
OY 262 KKVSYLSKQEK-----SLPERSNVGEIEMKFNFTISVSKCNTKLSYDLHESKLOHQLAVD 317
Db 391 RRCYFQKRELRKYNINISQSCNECLEANF---TLTKGCVK-----FSS 432
OY 318 LDRNINIQOKYCYPLIGSMT---YSVKEMEYLTR-----AIDRTGE 357
Db 433 MPRNVNM-----PVCGDASLACYQAQADELLREFYQGLVNAVCENTGGE 476

RESULT 13
OYVS151 PRELIMINARY: PRT: 1507 AA.
AC OYVS151;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE BRCA2 (FRAGMENT).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxId=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC Bignell G.R., Micklem G., Stratton M.R., Ashworth A., Wooster R.;
RA Submitted (Jul-1996) to the EMBL/Genbank/DBJ databases.
RL EMBL: 275664; CAA9994.1; -.
DR InterPro: IPR002093; -.
DR Pfam: PF00634; BRCA2_repeat; 7.

```


DR PROSITE: PS50138; BRCA2_REPEAT; 6.
 FT NON_TER 1 1507 1507
 SEQUENCE 1507 AA; 168659 MW; 2C78ABA31B93E257 CRC64;

Query Match 4.0%; Score 99.5; DB 6; Length 1507;
 Best Local Similarity 20.6%; Pred. No. 39;
 Matches 77; Conservative 55; Mismatches 143; Indels 99; Gaps 18;

OY 22 IIFVFONSTKYWALNLSLHYNNSTKSLFEPKTLISLK-----PLTEREL-70
 DB 436 IISGFVONSTYV-----SDSESGHTAPPLTLKQDFDSNRNLTPQKAETI 481
 OY 71 RIKEITEKLDQOIPPRP---THV--NTTSTATSTNTINPDYTCRGDOLHLLLEVR-124
 DB 482 ELSTILLESGSFETQPRKPSHIIOKNFEPENQOLIINSTSKMKQDDLLHTAPG 541
 OY 125 -DHLGRKKO---YGGDLRARMSSPALMAGASGKVTDFNNGTYLVSFTLLFWEQO-VLSLT-179
 DB 542 ISQVDSKKEGIIQKQKFACTLSRTSCNHSASGYSTDKN---VEFRGFYSARGTKLVN 597
 OY 180 LLIHPEGVSAIWSAR-----NOGYDRVIFTGQFVNGTSQV-----PLTERE-215
 DB 598 -----GSEALQAKKLFSDLEINIEETSVYDRSFSKINDYSMTQIEDCNKMLN 650
 OY 216 --HSECGLLNTNAELC-----QYLDNRQDSGFYCVRPQHPCAALT--HMYSKN 261
 DB 651 EPNKRCRLILONNIEMTDIFVEEYTESYRRTNEGNOCTDAGRNTCSNSESDESDSKN 710
 OY 262 KKVSYLSKQKESL--FERSNVGVEIMEK---NTIIVSKNTILKSLVHESKLOH---312
 DB 711 DTV-YIHEENGCLPCIDOHNDLKLPSQPMKEGNTQIKESLSDLCLEVMKAEETSHVTK 769
 OY 313 ---OLAVDLQNI 322
 DB 770 SNKQOLANTGONI 783

RESULT 14
 ID 09STU3 PRELIMINARY; PRT; 795 AA.
 AC 09STU3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE HYPOTHETICAL 90.2 KDA PROTEIN.
 GN 12307.30.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Cholsene N., Robert C., Bottler P., Wincker P., Cattolico L.,
 RA Attiguenave F., Saurin W., Weissenbach J., Mewes H.W., Mayer K.F.X.,
 RA Lemcke K., Schueller C., Queller F., Salanoubat M.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF049746; CAB41853.1;
 KW Hypothetical protein.
 SO SEQUENCE 795 AA; 90233 MW; 3E4A3EBD36105088 CRC64;

Query Match 4.0%; Score 98.5; DB 10; Length 795;
 Best Local Similarity 19.1%; Pred. No. 18;
 Matches 100; Conservative 79; Mismatches 181; Indels 163; Gaps 26;
 OY 33 VMSALNLSLHYNNSTKSLFEPKTLISLKPLTERE-----LRKEITEKLDQO 82

DB 149 VMSLKNL-----RTSRSGFEVNLAKITLTTEILLISVAKHRRMAKLVANVH 205
 OY 83 I-----PPRFTHVNTTSAT--HSTATILNPROVYVWQ 119
 DB 204 VDRALAMMPQALADYRALLSLRWPPQ---LSTLSASLDSKSNVQNT-----259
 OY 117 LHLLEVRDHLGRKKOYGGDLRARMSSPALMAGASGKVTDFNNGTYLVSFTLLFWEQO-VLSLT-179
 DB 251 ---LFNMEGSL--KSOYCGSF--HALCSLOGIOLQKRSKROGLIKHBNLPHQPLMAHEL 404
 OY 177 LSTLLIHPEGVSAIWSARNOGYDRVIF--TQYVNGTSVISECGLLINNAEL-VYLT-244
 DB 305 VNPETV-ASQRRFTKWEKPEFIFALVKITRDVDSMDL-----LPIYV 460
 OY 235 NNDQEGFVCRPQHPCAALTHV-----SKKKVSYT-----497
 DB 351 EAKLAGYSGREWVSAMVSSLSLYLVEKFIPIYVQDLEANTEDLRSAKYSWHLIDLM 410
 OY 268 ---SKQESLPERSNVGVLEIMEKFNITSVK---CNTLKSVALHESKRLHCLA-----415
 DB 411 ISFDKRVSLVSQGI--LSLQEDGNLRTLSLSEYCDRDMDLMAELTELDERLVRKEH 449
 OY 316 VLDLRNINIQWK-----YCPPLISMTYSVKEKYLITATIRISEKNIVYSLR 477
 DB 470 IINDRNMTAKVQDELISSNVYRPPIISIF-----LOHLSSTIERS---KSPVALYK 520
 OY 368 QHRPEPDIQFIRBALN--VKAIGHLLRSFOT--WIIKTEN;KEMNDAHRSYHRY 424
 DB 521 AR-----FLRIASPTIHKFLQCLLRQDADGLTALTEN-----NOLIKVSNINA 577
 OY 425 IOTLIKIFODLSVSIIDAMDITAY---GTNNVHPQHVW 464
 DB 568 GHYI-----ESYLEWSEDEVFLEMOTGQ-HDQPEVP 539

RESULT 15
 ID 096167 PRELIMINARY; PRT; 1004 AA.
 AC 096167;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE SERA ANTIGEN/PAPAIN-LIKE PROTEASE WITH ACTIVE SER.
 GN PFB0360C.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99021743; PubMed=9804551;
 RA Gardner M.J., Jettellin H., Carucci D.J., Cummings L.M., Atayind L.,
 RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Peterson J.,
 RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Petras M.,
 RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
 RA Fraser G.M., Adams M.D., Venter J.C., Hoffman S.L.;
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum".
 RL Science 282:1126-1133(1998).
 DR EMBL: AF001389; AAC71861.1;
 DR InterPro: IPR000169;
 DR InterPro: IPR000668;
 DR Pfam: PF00112; Peptidase_C1; 1.
 DR PRINTS: PR00705; PAPA1N.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.
 KW Hydrolase; Protease; Thiol protease.
 SO SEQUENCE 1004 AA; 116573 MW; 1F1750B2A9BE10C9 CRC64;

Query Match 4.0%; Score 98.5; DB 5; Length 1004;
 Best Local Similarity 17.5%; Pred. No. 26;
 Matches 85; Conservative 88; Mismatches 151; Indels 191; Gaps 22;

```

OY 23 IFTVONSTKWSALNLSISLHYMNSTKSLFPRKPLISKLPTETELRIKETIEXLDOQ 82
Db 71 IHASPSOTKEMSPNNLT-----SKKKKKKEIRPKDIMSNSDS- 109
OY 84 IPRPPTHVNTTTSATHTATILNPRDTCRGDQHLILEYDHLGRKQYGGDFLRAM 142
Db 110 -----SNTSSINKQN---NNQIKSVL-LKENKG-----VKI 136
OY 143 SSPALMAGASGYTDFNNGTYLVSTLFMEGQVSLSLTI-HPSEGVSAIAMSARNOGYDR 201
Db 147 TQPC-----NVNLSIFLVPHIYIDVETRYNNIELKTEL 169
OY 202 VFTG--QFVNGTSQVHSECGLIINTNAELQYLDNRQDEGFCYRPOHMPCAALTHMYS 259
Db 170 DEFSSIKFKDOTTLELRTSDOTLMTNPNVGSRDITLQKRLYN-----CA----- 216
OY 210 KKKKVSYSKQEKSLFERSNNGVEIMEKFWTISVSKCNTLSVDLHESGKLOHOLAVDLD 319
Db 217 EKKTFKFW-----VYIKDNI-----LTLK-WKVEYETGVYNNK--VDIR 251
OY 320 RNINIQWQKXCPPLIGSMYTSYKEMEYLTPRAIDRTGKENTVIIVISLQHFRRFPIDVFI 379
Db 254 Q--YKKKELTPITTIQIHSYSENKO-THLE---SKNYVITKIDIPETCDVMATNCEL 303
OY 380 RRALNVHKAIO-HLLRSPT---MWIKTENREMYN-----DAERSDPH----- 422
Db 404 SGNINIEKLECTLLVONNDTSSECFYVSNDYRENFNOIKAEADDEDFRNYHLDITIN 363
OY 423 -----GYIOYLIIKDIPODLVSIIIDAWDITIAVGTNNVHPQHYVGNQINI 469
Db 364 NILKRIYKINKNKGKKELITLELDONFLKESITDYCKILREIDTNGT-LVNHELGNNDV 422
OY 470 LLNYI 474
Db 423 FNNLI 427

```

Search completed: October 12, 2001, 16:05:18
Job time: 161 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2001, 16:03:25 ; Search time 34.67 Seconds
(without alignments)
956,485 Million cell updates/sec

Title: US-09-729-454-2

Perfect score: 2908

Sequence: 1 MMSNMLQKTLILISFSVY.....IHPDHVIGNQIMFLNIEIC 547

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum hit seq length: 0
Maximum hit seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A-Geneseq-0601.*
1: /SIDSB/gcqdala/geneseq/geneseq/AA1980.DAT.*
2: /SIDSB/gcqdala/geneseq/geneseq/AA1981.DAT.*
3: /SIDSB/gcqdala/geneseq/geneseq/AA1982.DAT.*
4: /SIDSB/gcqdala/geneseq/geneseq/AA1983.DAT.*
5: /SIDSB/gcqdala/geneseq/geneseq/AA1984.DAT.*
6: /SIDSB/gcqdala/geneseq/geneseq/AA1985.DAT.*
7: /SIDSB/gcqdala/geneseq/geneseq/AA1986.DAT.*
8: /SIDSB/gcqdala/geneseq/geneseq/AA1987.DAT.*
9: /SIDSB/gcqdala/geneseq/geneseq/AA1988.DAT.*
10: /SIDSB/gcqdala/geneseq/geneseq/AA1989.DAT.*
11: /SIDSB/gcqdala/geneseq/geneseq/AA1990.DAT.*
12: /SIDSB/gcqdala/geneseq/geneseq/AA1991.DAT.*
13: /SIDSB/gcqdala/geneseq/geneseq/AA1992.DAT.*
14: /SIDSB/gcqdala/geneseq/geneseq/AA1993.DAT.*
15: /SIDSB/gcqdala/geneseq/geneseq/AA1994.DAT.*
16: /SIDSB/gcqdala/geneseq/geneseq/AA1995.DAT.*
17: /SIDSB/gcqdala/geneseq/geneseq/AA1996.DAT.*
18: /SIDSB/gcqdala/geneseq/geneseq/AA1997.DAT.*
19: /SIDSB/gcqdala/geneseq/geneseq/AA1998.DAT.*
20: /SIDSB/gcqdala/geneseq/geneseq/AA1999.DAT.*
21: /SIDSB/gcqdala/geneseq/geneseq/AA2000.DAT.*
22: /SIDSB/gcqdala/geneseq/geneseq/AA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	103.5	3.6	1237	21	AAV96999
2	101	3.5	452	21	AAV70502
3	97	3.3	252	21	AAAB01392
4	97	3.3	252	21	AAV99401
5	97	3.3	252	22	AAAB87571
6	97	3.3	252	22	AAAB6150
7	96.5	3.3	471	21	AAV70274
8	96.5	3.3	1714	21	AAAB18275
9	94.5	3.2	379	21	AAAG12775
10	94.5	3.2	383	21	AAAG12774
11	94.5	3.2	395	22	AAAB5799

12	93	3.2	404	21	AAV70275	Human purino recep
13	93	3.2	497	21	AAV70277	Human purino recep
14	91.5	3.1	592	11	AAAR05227	Sequence of rat C
15	90	3.1	253	21	AAAB43066	Human ORFX ORF2830
16	89	3.1	1792	21	AAAB48447	Mouse lamrin 8 po
17	89	3.1	1816	21	AAAB48446	Mouse lamrin 8 po
18	88.5	3.0	905	22	AAAB72198	Pepper Bsz amino a
19	88.5	3.0	1004	21	AAAB18214	Plasmodium falcipa
20	88.5	3.0	2013	21	AAAB18265	Plasmodium falcipa
21	88.5	3.0	2954	20	AAV01632	Amino acid sequenc
22	88	3.0	839	21	AAAB56864	Human prostate can
23	87.5	3.0	739	15	AAAR49733	Sequence of the 1E
24	87	3.0	727	21	AAAG54101	Arabidopsis thalia
25	87	3.0	765	21	AAAG54100	Arabidopsis thalia
26	87	3.0	815	21	AAAG54099	Arabidopsis thalia
27	87	3.0	986	18	AAAB09930	Arabidopsis thalia
28	87	3.0	986	18	AAAB23552	SEQ ID 1 from W097
29	87	3.0	1164	18	AAW22469	Cysteine loop prot
30	87	3.0	1164	21	AAAB01263	Streptococcal C4a
31	86.5	3.0	739	11	AAAR08118	SCP49 peptidase (
32	86.5	3.0	739	12	AAAR10316	Vascular cell adhe
33	86.5	3.0	739	12	AAAR13906	Human VCAM-1. Hom
34	86.5	3.0	739	14	AAAR38548	Human VCAM-1. Synthe
35	86.5	3.0	739	19	AAAB46734	Amino acid sequenc
36	86.5	3.0	744	21	AAV94408	Human VCAM-1/19G1-
37	86.5	3.0	1073	18	AAW32063	Human ST receptor
38	86.5	3.0	1073	19	AAW37371	Human ST receptor
39	86	3.0	355	21	AAV70408	C. elegans alpha-1
40	85.5	2.9	447	21	AAV70276	Human purino recep
41	85.5	2.9	739	13	AAAR21081	Sucrose synthase P
42	85.5	2.9	805	21	AAAB21841	VCAM variant with
43	85.5	2.9	805	21	AAAB16282	Human purino recep
44	85.5	2.9	805	21	AAAB16336	Eucalyptus grandis
45	85.5	2.9	3079	15	AAAR39926	Eucalyptus grandis

ALIGNMENTS

RESULT 1	
AAV96999	standard; Protein; 1237 AA.
ID	AAV96999
XX	AAV96999;
AC	31-OCT-2000 (first entry)
DT	31-OCT-2000 (first entry)
XX	S. cerevisiae essential gene YGR090W product.
DE	S. cerevisiae essential gene YGR090W product.
XX	YGR090W; essential gene; survival; growth; antifungal; inhibitor.
KW	Saccharomyces cerevisiae.
OS	Saccharomyces cerevisiae.
XX	WO200039342-A2.
PN	06-JUL-2000.
XX	06-JUL-2000.
PD	29-DEC-1999; 99MO-US31042.
XX	29-DEC-1999; 99MO-US31042.
PF	31-DEC-1998; 98US-0114734.
XX	31-DEC-1998; 98US-0114734.
PR	(MILL-) MILLENNIUM PHARM INC.
PA	Bulawa C;
XX	Bulawa C;
PI	WPI: 2000-452421/39.
XX	WPI: 2000-452421/39.
DR	N-PSDB; AAAS1777.
XX	N-PSDB; AAAS1777.
DR	Use of essential Saccharomyces cerevisiae genes and polypeptides for
XX	identifying antifungal compounds useful for treating fungal infections
PT	Claim 36; Fig 3; 117pp; English.
XX	Claim 36; Fig 3; 117pp; English.
PS	
XX	

CC The Saccharomyces cerevisiae proteins shown in AAY96997-97013 are
CC encoded by genes essential for survival of Saccharomyces. Essential genes
CC are required for growth (e.g. metabolism, division or reproduction) and
CC survival of an organism. These essential genes can be used to identify
CC therapeutic antifungal agents. The therapeutic agents can reduce or
CC prevent growth, or decrease pathogenicity or virulence, and preferably,
CC kill the fungi. The agents, fungal growth inhibitors (which may inhibit
CC transcription or translation of the essential gene) are useful for
CC creating an organism, preferably human with a fungal infection
CC (claimed).

5Q Sequence 1237 AA;

Query Match 3.6%; Score 103.5; DB 21; Length 1237;
Best Local Similarity 19.3%; Pred. No. 0.5;
Matches 102; Conservative 87; Mismatches 175; Gaps 28

[illegible]

RESULT 2

ID AAY70502 standard; Protein; 452 AA.

AY70502 standard; 452 AA.

AC MAY 7 05 02;

DT 04-JUL-2000 (first entry)

Corn farnesyltransferase beta subunit.

Corn; farnesyltransferase beta subunit; transgenic plant;

drought tolerance; cell growth.

05 Zee mayb.

PN WO200014207-A2.

PD 16-MAR-2000.

XX 07-SEP-1999; 99WO-US20419
PF

PR 08-SEP-1998; 9805-0099521.

PA (DUPO) DU PONT DE NEMOURS & CO E I.

PI Cahoon RE, Miao G, Powell W;

DR WPI; 2000-256964/22.

DR N-PSDB; AA251791.

PT New isolated polynucleotide encoding farnesyltransferase polypeptide is
PT useful for producing transgenic plants with an altered level of
PT farnesyltransferase -

PS Claim 14; Page 39-40; 51pp; English.

CC The present sequence is a farnesyltransferase beta subunit from
CC clone P0127.cmbu.14r isolated from corn nucellus tissue cDNA library
CC The coding sequence of this protein is used in the construction
CC of a chimeric gene to produce transgenic plants with altered level
CC of farnesyltransferase subunit. Plants with decreased farnesyltransferase
CC activity may have enhanced tolerance to drought stress. Nucleic acids
CC encoding all or a part of farnesyltransferase proteins can be used in
CC studies to understand cell growth in plants and provide genetic tools
CC to control cell growth and improve tolerance to drought in mature plants

sq Sequence 452 AA;

Query Match 3.58; Score 101; DB 21; Length 452;
Best Local Similarity 20.18; Pred. No. 0.17;
Matches 60; Conservative 55; Mismatches 100; Indels 84; Gaps 15

QY	86	PEPFENHTVTSATSTATIIINPR--DTGCRDQDILILEVHNLJAGKQVCGCTLAAM	14
Db	132	pqgplhatltyaaavlvtvltsgeralslnlgnlplnmlqmkdvsairimhggelwtra	191
QY	144	SSPALTAGASGAKMDF---NNGTLYVSTFLWESQVS-----ESTLLI	189
Db	192	sytlasvsaivnlldlklakvgvdyatrcqy-egslasepyaahnyrttqalaalll	250
QY	184	HSEGSASL---WRARNOGDKIIFFGK-----FNGHSVFEECLTINSNAE	229
Db	251	neaekvdpilpvgwaftqgye-cgfgtrtnklvqgcyftqgaalatrf	300
QY	230	LCVELDRDDEATYCCKAQ-HMPEALITVTRNREYSTLQENSLHRSKVGVEKKD	288
Db	301	lllvdqkllkssysckrfpsgedactssyqct-----akssasavdyakfdrf	349
QY	289	RKHIDVTRCKREKIEFFCOVGKMPVPGGYTLQCKWITTFQWVQLDTKIKNGTEGK	347
Db	350	-----lqgnqngq---plfnlaldq-qlv-llscqy-----lqndldk	384
RESULT 3			
ID	AA01392	standard: Protein: 252 AA.	
AC	AA01392:		
XX	20-OCT-2000 (first entry)		
DE	Neuron-associated protein.		
XX	Neuron associated protein: NEUP; neurological disorder; epilepsy;		
KM	ischemic cerebrovascular disease; stroke; cerebral neoplasm;		
KM	Alzheimer's disease; Pick's disease; Huntington's disease;		
KM	dementia; Parkinson's disease; demyelinating disease; meningitis;		
KM	prior disease; kuru; Creutzfeldt-Jakob disease; neurofibromatosis;		
KM	cerebral palsy; muscular dystrophy; central nervous system; CNS;		
KM	peripheral nervous system; PNS; myopathy; schizophrenia;		

RESULT 3

ID AAB01392 standard; Protein; 252 AA.

ID	standard; Protein; 252 AA.
AAB01392	

AC AAB01392;

DT 20-OCT-2000 (first entry)

Neuron-associated protein.

KW Neuron associated protein; NEUP, neurological disorder; epilepsy;

KW Ischemic cerebrovascular disease; stroke; cerebral neoplasm;

KW Alzheimer's disease; Pick's disease; Huntington's disease;

dementia; Parkinson's disease; demyelinating disease; meningitis

Prion disease; kuru; Creutzfeldt-Jakob disease; neuritic fibrin test

cerebral palsy; muscular dystrophy; central nervous system; CNS;

KW peripheral nervous system; PNS; myopathy; schizophrenia;

KM actinic keratosis; arteriosclerosis; atherosclerosis; bursitis;
 KM cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
 KM myelofibrosis; paroxysmal nocturnal haemoglobinuria; cancer;
 KM autoimmune disease; inflammation; acquired immunodeficiency syndrome;
 KM AIDS; Addison's disease; adult respiratory distress syndrome;
 KM allergy; ankylosing spondylitis; amyloidosis; anaemia; asthma;
 KM Werner syndrome; trauma; human.
 XX
 US Homo sapiens.
 XX
 EH Key
 FT Peptide
 FT 1..22 Location/Qualifiers
 FT 10..54 label= Signal peptide
 FT Modified-site
 FT 62 /note= "Potential phosphorylation site"
 FT Modified-site
 FT 93 /note= "Potential glycosylation site"
 FT Modified-site
 FT 119 /note= "Potential phosphorylation site"
 FT Modified-site
 FT 127 /note= "Potential phosphorylation site"
 FT Modified-site
 FT 137 /note= "Potential glycosylation site"
 FT Modified-site
 FT 143 /note= "Potential glycosylation site"
 FT Modified-site
 FT 189 /note= "Potential glycosylation site"
 FT Modified-site
 FT 242 /note= "Potential phosphorylation site"
 FT Modified-site
 FT /note= "Potential phosphorylation site"
 XX MO200034477-A2.
 XX 15-JUN-2000.
 XX
 PE 10-DEC-1999; 99WO-US30408.
 XX
 PK 11-DEC-1998; 98US-0210083.
 PK 11-DEC-1998; 98US-9123456.
 PR 09-FEB-1999; 99US-0119365.
 PR 16-MAR-1999; 99US-0124687.
 XX
 PA (INCYTE) INCYTE PHARM INC.
 XX
 PI Tang YT, Yue H, Baughn MR, Hillman JL, Lai P, Au-Young J, Yang J;
 PI Lu DM, Azimzal Y;
 DR MPI: 2000-423423/36.
 DR N-PSDB; AAA47433.
 XX
 PT New human neuron-associated proteins and polynucleotides encoding them,
 PT useful for diagnosis, treatment and prevention of cell proliferative
 PT disorders including cancer, neuronal and neurological disorders
 XX
 PS Claim 1: Page 117; 145pp; English.
 XX
 XX Human neuron-associated proteins (NEUAP) can be used for
 CC treating or preventing a disorder associated with decreased
 CC expression or activity of NEUAP. Antagonists of NEUAP are useful for
 CC treating or preventing a disorder associated with increased expression
 CC or activity of NEUAP. NEUAP or their fragments or derivatives are
 CC useful for treating neurological disorder such as epilepsy, ischemic
 CC cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's
 CC disease, Pick's disease, Huntington's disease, dementia and
 CC Parkinson's disease. NEUAPs are also useful for treating other
 CC demyelinating diseases, bacterial and viral meningitis, prion
 CC diseases including kuru, Creutzfeldt-Jacob disease, nutritional and
 CC metabolic diseases of the nervous system, neurofibromatosis, other
 CC developmental disorders of the central nervous system, cerebral
 CC palsy, neuromuscular disorders, autonomic nervous system disorders,
 CC cranial nerve disorders, spinal cord diseases, muscular dystrophy and
 CC other neuromuscular disorders, peripheral nervous system disorders.

CC inherited, metabolic, endocrine, and toxic myopathies, mental
 CC disorders including mood, anxiety and schizophrenic disorders, a cell
 CC proliferative disorder such as actinic keratosis, arteriosclerosis,
 CC atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective
 CC tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
 CC haemoglobinuria, cancers of the adrenal gland, bladder, bone,
 CC bone marrow, brain, breast, cervix, and an autoimmune/inflammatory
 CC disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
 CC disease, adult respiratory distress syndrome, allergies, ankylosing
 CC spondylitis, amyloidosis, anemia, asthma, Werner syndrome,
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic
 CC infections, and trauma. This protein was given the Incyte ID no.
 CC 3070147CD1.
 CC
 XX
 SO Sequence 252 AA;

Query Match 3.38; Score 97; DB 21; Length 252;
 Best Local Similarity 23.38; Pred. No. 0.17; Mismatches 28; Gaps 6;
 Matches 35; Conservative 32; Indels 55;

QY 131 RKQYG-GDFLRMSSPALTAGAGKVDFFNNGTYLVFTLPFEGQVSLILLHPSEG-188
 Db 97 kklfgvgdly-snlkvalhllvtgklvdhngvgfshqdnatqgnlslstlypskav 155
 QY 189 -----NSALMRNRNGYDKLIIFGKRVNGTSHVTEGGLTNSNAELCEYLD 235
 Db 156 efngeqgflieakaskklfncme-wek-vergr-----rtslcthpklcsrdh 203
 QY 236 DRDQEAFCMKPQHMPCEALTYMTTRNREV 265
 Db 204 agssatwscsqgfkvcvyafystdytlv 233

RESULT 4
 ID AAY99401 standard; Protein: 252 AA.
 XX
 AC AAY99401;
 XX
 DT 08-AUG-2000 (first entry)
 XX
 DE Human PRO1327 (UNQ687) amino acid sequence SEQ ID NO:218.
 XX
 KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
 KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
 OS Homo sapiens.
 XX
 PN MO200012708-A2.
 XX
 PD 09-MAR-2000.
 XX
 PF 01-SEP-1999; 99WO-US20111.
 XX
 PR 01-SEP-1998; 98US-0098716.
 PR 01-SEP-1998; 98US-0098749.
 PR 01-SEP-1998; 98US-0098750.
 PR 02-SEP-1998; 98US-0098803.
 PR 02-SEP-1998; 98US-0098821.
 PR 02-SEP-1998; 98US-0098843.
 PR 09-SEP-1998; 98US-0099536.
 PR 09-SEP-1998; 98US-0099596.
 PR 09-SEP-1998; 98US-0099599.
 PR 09-SEP-1998; 98US-0099602.
 PR 09-SEP-1998; 98US-0099741.
 PR 10-SEP-1998; 98US-0099754.
 PR 10-SEP-1998; 98US-0099763.
 PR 10-SEP-1998; 98US-0099792.
 PR 10-SEP-1998; 98US-0099808.
 PR 10-SEP-1998; 98US-0099812.

DB 204 agsawtscspkfvvcvyaitystidrlv 233

RESULT 5

AAB87571
ID AAB87571 standard; Protein: 252 AA.

XX AAB87571:

XX 15-MAY-2001 (first entry)

XX Human PRO1327.

XX Human; PRO protein; mapping.

XX Homo sapiens.

XX WO200116318-A2.

XX 04-MAR-2001.

XX 24-AUG-2000; 2000MO-US23328.

XX 01-SEP-1999; 99MO-US20111.

XX 15-SEP-1999; 99MO-US21090.

XX 07-DEC-1999; 99US-0169495.

XX 04-DEC-1999; 99US-0170262.

XX 11-JAN-2000; 2000US-0175481.

XX 18-FEB-2000; 2000MO-US04342.

XX 22-FEB-2000; 2000MO-US04414.

XX 01-MAR-2000; 2000MO-US05601.

XX 03-MAR-2000; 2000US-0187202.

XX 25-APR-2000; 2000US-0199397.

XX 22-MAY-2000; 2000MO-US14042.

XX 05-JUN-2000; 2000US-0209812.

XX (GETH) GENENTECH INC.

XX Eaton PL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;

XX Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;

XX WPI: 2001-183260/18.

XX N-PsDM; AAP92103.

XX Eighty four nucleic acids encoding PRO polypeptides, useful in

XX molecular biology, including use as hybridization probes, and in

XX chromosome and gene mapping.

XX Claim 12: Fig 92: 278pp; English.

XX The present sequence is a human PRO polypeptide (secreted and

XX transmembrane). The PRO protein, and PRO agonists, PRO antagonists or

XX anti-PRO antibodies are useful for preparation of a medicament useful in

XX the treatment of a condition which is responsive to the PRO protein,

XX agonists, antagonists or anti-PRO antibodies. The PRO protein may also be

XX employed as molecular weight markers for protein electrophoresis. The PRO

XX coding sequence has applications in molecular biology, including use as

XX hybridisation probes, and in chromosome and gene mapping.

XX Sequence 252 AA:

XX Query Match 3.38; Score 97; DB 22; Length 252;

XX Best Local Similarity 23.38; Pred. No. 0.17; Mismatches 55; Indels 28; Gaps 6;

XX Matches 15; Conservative 32; Mismatches 55; Indels 28; Gaps 6;

DB 131 RKQYG-GDFLRARMSPPALTAGASGVMDFNNGTYLVFTLFWEGCVSLILLHPSEG-188

DB 97 KLIIGVQDLY-SNKTVALNLLVTKIVGHGTFVHFQHNATGQGNLSLTPPKAV-155

DB 189 -----ASALMRARNGYDKIIFPKGKRVNNTSHVFTECGTLTNSAELCEYLD-235

DB 156 efhqeqqifiekasklifncme-wek-vergr-----rslcthpakicsrdh 203

OY 236 DRDQEAFCMKPKQHMPCEALTYMTTRNEV 265

DB 204 agsawtscspkfvvcvyaitystidrlv 233

RESULT 6

AAB6150
ID AAB6150 standard; protein: 252 AA.

XX AAB6150:

XX 02-APR-2001 (first entry)

XX Protein of the invention #62.

XX Secreted; transmembrane; gene therapy.

XX Unidentified.

XX WO200078961-A1.

XX 28-DEC-2000.

XX 18-FEB-2000; 2000MO-US04342.

XX 23-JUN-1999; 99US-0141037.

XX 20-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 01-SEP-1999; 99MO-US20111.

XX 29-OCT-1999; 99US-0162306.

XX 30-NOV-1999; 99MO-US28313.

XX 02-DEC-1999; 99MO-US28551.

XX 16-DEC-1999; 99MO-US30095.

XX 05-JAN-2000; 2000MO-US00219.

XX 06-JAN-2000; 2000MO-US00376.

XX (GETH) GENENTECH INC.

XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

XX Watanabe CK, Williams PM, Wood WI;

XX WPI: 2001-071395/08.

XX Secreted and transmembrane proteins and nucleic acids designated PRO,

XX useful as hybridization probes, in chromosome and gene mapping and gene

XX therapy.

XX Claim 1; Fig 124: 787pp; English.

XX The present invention relates to secreted and transmembrane proteins,

XX probes, in chromosome and gene mapping and in the generation of

XX anti-sense RNA and DNA. They may also be used used to generate either

XX transgenic animals or knockout animals which are in turn useful for

XX development and screening of therapeutically useful reagents.

XX The nucleic acids may also be used in gene therapy.

XX Sequence 252 AA:

XX Query Match 3.38; Score 97; DB 22; Length 252;

XX Best Local Similarity 23.38; Pred. No. 0.17; Mismatches 55; Indels 28; Gaps 6;

XX Matches 35; Conservative 32; Mismatches 55; Indels 28; Gaps 6;

DB 131 RKQYG-GDFLRARMSPPALTAGASGVMDFNNGTYLVFTLFWEGCVSLILLHPSEG-188

DB 97 KLIIGVQDLY-SNKTVALNLLVTKIVGHGTFVHFQHNATGQGNLSLTPPKAV-155

DB 189 -----ASALMRARNGYDKIIFPKGKRVNNTSHVFTECGTLTNSAELCEYLD-235

CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.

SU Sequence 1714 AA:

Query Match 33%; Score 95; DB 21; Length 1714;
Best Local Similarity 19.8%; Pred. No. 6.6; Mismatches 112; Indels 82; Gaps 15;
Matches 61; Conservative 53;

QY 225 NSMAELCEYLDDRCDFAYC---MKPDMPCFALTYMTTRNREVS-----L 268
DB 1107 nmndlnhdndnngfnsceelnqprveqkcegekyedkknkylynnnyklogknei 1166
QY 269 TDKENSIFHRSKYGVEM--MKDRKHIDVTMCNKRKEIETQVG--MKPPYEGITLQCK 324
DB 1167 hddynlksbysrtnynlfnlkdnh-----nmndgkscelkkslpyvkekynlen- 1219
QY 325 WITFCNQVQLTTRKINGCLKGLLYLDSDTLRQWYFFKRVYKTLKFP----DLHET 379
DB 1220 -----tlyeflg-----llygdksgyykciimnkvkryamkvylkecneifvd 1263
QY 180 GIFKKHLLDAERHTQIQMKKHSYFVTQYSLI-DHDIYPREIDRLSGDKNTAVITF 438
DB 1264 ntlkkyllfknpkhl-----lsllydftcmnylicimdyceg---stllidy 1308
QY 439 GQHFREPPIDFIRRAIGVOKAIEFLRSP---ATKYI---IKTEMI-----REWHI 485
DB 1309 fmslvpasldvy-----elkkimknlfialdfhsmnlhndklenlmkknkkkkrf 1363
UY 486 ETFFCQF 493
DB 1364 nyekygsf 1371

RESULT 9

AA012775
ID AA012775 standard; Protein; 379 AA.

XX
AC AA012775;

XX
DT 17-OCT-2000 (first entry)

XX
FE Arabidopsis thaliana protein fragment SEQ ID NO: 12018.

XX
KM Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

XX
OS Arabidopsis thaliana.

XX
PN EPI03405-A2.

XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.

XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 21-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.

PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132487.
PR 11-MAY-1999; 990S-0132863.
PR 14-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 16-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139412.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139859.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144315.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.

```

PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147933.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149172.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159325.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

```

```

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match

3.2% Score 94.5; DB 21; Length 379;

Best Local Similarity 20.5%; Pred. No. 0.61;

Matches 53; Conservative 31; Mismatches 72; Indels 103; Gaps 10;

```

QY 299 KREKIETC--QVG-----MKPPVGGYTLGCKWITTFNQ 342
      : : : : : : : : : : : : : : : : : : : : : :
DB 43 rrrreerdcysgkwtfdedyplydsscaylssalsqngndfsyqkwtwipkcs1 102
      : : : : : : : : : : : : : : : : : : : : : :
QY 333 VOLPTIKNGCKLTYLIDSTLR-QM-----IYFFPVVKIKF 373
      : : : : : : : : : : : : : : : : : : : : : :
DB 103 prfialkfllgkmqgkrlmlyvgdsnmrnqeslvcjvgsvlptlkklltvgp----tms1 158
      : : : : : : : : : : : : : : : : : : : : : :
QY 374 FDLH-----ETGIFKKHL-----LDAFRHTQJWKRSY 404
      : : : : : : : : : : : : : : : : : : : : : :
DB 159 hslfdetsiefcwapllvelkryvdkrvlhdslednarykwrvdvlrfsahwrlsq 218
      : : : : : : : : : : : : : : : : : : : : : :
QY 404 PEVFFQLYSLIDHYIPREIDRLSGCKNTAIVITGQHPRI:ITIRKALGVKALEP 464
      : : : : : : : : : : : : : : : : : : : : : :
DB 219 rwswdwy--mdgnkflkamd-----fmvayevqlltwkkyv- 254
      : : : : : : : : : : : : : : : : : : : : : :
QY 464 LFLSPATKVIITKTNIRE 482
      : : : : : : : : : : : : : : : : : : : : : :
DB 255 inldpsktkvlftrvspre 273
      : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
AAG12774
ID AAG12774 standard: Protein; 383 AA.
XX AC AAG12774;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 12017.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PE 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.

```

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 25-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137558.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 04-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-01452913.
PR 27-JUL-1999; 99US-01452918.
PR 27-JUL-1999; 99US-01452919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149476.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155569.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

XX	14-DEC-2000.
XX	
XX	02-JUN-2000; 2000MO-N200086.
XX	
XX	04-JUN-1999; 99US-0325932.
XX	
XX	(GENE-) GENESIS RES & DEV CORP LTD.
XX	(FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX	
XX	Flinn B, Lasham A;
XX	
XX	WPI: 2001-061724/07.
XX	
XX	N-PSDB; AAF44826.
XX	
XX	Novel defender against cell death polynucleotide useful for modulating
XX	PT programmed cell death pathway and specific development pathways in
XX	PT forestry plant -
XX	
XX	Claim 22: Pages 125-126; 142pp: English.
XX	
XX	The present invention relates to coding sequences (see AAF44740-F44840
XX	and AAF44843-F44844) and proteins (see AAB55714-B55814) involved in
XX	CC programmed cell death (PCD; apoptosis). The coding sequences and proteins
XX	CC of the present invention are useful for modulating a part of cell death
XX	CC pathway and various developmental pathways in a forestry plant, by
XX	CC stably incorporating one of the present coding sequences into the genome
XX	CC of the forestry plant, where the coding sequence provides a PCD pathway
XX	CC that is not present in a native form of the forestry plant.
XX	
XX	Sequence 395 AA:

Query Match	3 28;	Score 93.5;	DB 22;	Length 395;
Best Local Similarity	24.08;	Pctd. No 0 83;	76;	Indels 51; Gaps 11
Matches	50;	Conservative	31;	Mismatches 51; Gaps 11
<hr/>				
QY	234	LDDEDDGATVCKRQHPNHCALYLTNTNREVSLLTQKENSFLHRSKVYEMMKTPKML	294	
Db	160	vddvddqkafldpqsvlnseelllcttdrvnlslkvenssly-----s	204	
<hr/>				
QY	294	VTNNKKEKIETQVCKMKP--PVPGATLQKKITTFPCNVCLDTIMNLYKRNLYL	461	
Db	205	lsglnkhsstelfcshafspafldpfksldkfl-dycnqfpl-sklktda-----	256	
<hr/>				
QY	352	LGDSTLQWLYFPRVVKVTKLFPDLHENGIRK-----KHLLDA-----ERRT	394	
Db	256	lyqkdlspvkwew-estlrapiadlndt--klsysdlnqeeakafldiacllrcchhda	312	
<hr/>				
QY	395	QIQ-VRKNHSY---PVTPLDYLSDLHND	417	
Db	313	alslmksqwtngtgnflndqsklvevd	340	

RESULT	12
AAV70275	
ID	AAV70275 standard; Protein: 404 AA.
XX	
AC	AAV70275;
XX	
DT	20-JUN-2000 (first entry)
XX	
DE	Human purino receptor P2X _{2b} (hP2X _{2bR}) protein.
XX	
KW	Purino receptor; human; hP2X _{2bR} ; ligand-gated ion channel; analgesic;
KW	auditory; agonist; antagonist; modulator; pain; treatment; vestibular;
KW	neuroendocrine disorder; gastrointestinal tract.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Misc-difference
FT	380..381
FT	Location/Qualifiers
FT	/note= "Encoded by a region of 200 nucleotides
FT	represented as base pairs 1146..1346 in AAV70275."

```

XX      W0200011167-A2.
XX      02-MAR-2000.
XX      20-AUG-1999; 99MO-US19042.
XX      20-AUG-1998; 98US-0137458.
XX      (ABBO ) ABBOTT LAB.
XX      Lynch KJ, Burgard EC, Metzger RE, Niforatos W, Touma EB;
XX      WPI: 2000-237649/20.
XX      N-PSDB: AA251322.
XX
XX      Novel DNA encoding human P2X2 receptor, vectors and methods of
XX      identifying modulators of the receptor, useful for relieving pain and
XX      treating auditory, vestibular, neuroendocrine and gastrointestinal
XX      tract disorders.
XX      Claim 21: Fig 8B; 45pp; English.
XX
XX      The present amino acid sequence is the human purino receptor P2X2
XX      (hP2X2) protein, derived from human pituitary gland tissues. P2X2
XX      receptors are ligand-gated ion channels, activated by ATP and other
XX      natural or synthetic nucleotides. It has analgesic and auditory activity.
XX      P2X2 sequences are useful for recombinant production and identification
XX      of therapeutic compounds, that modulate or interact with the P2X2
XX      receptor. P2X2 agonists and antagonists are useful for relieving pain
XX      and treating neuroendocrine, auditory and vestibular, and
XX      gastrointestinal tract disorders.
XX
XX      Sequence 404 AA:
XX
XX      Query Match 3.2%; Score 93; DB 21; Length 404;
XX      Best Local Similarity 22.9%; Pred. No. 0.98;
XX      Matches 69; Conservative 36; Mismatches 86; Indels 110; Gaps 18;
XX
XX      11 LLIISFSVVTMMIFISQNFKLSALNLSISVHYMNSAKSLFPKTSLI-PLKPLTET 69
XX      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      50 llllyl---vwyvllvqksyge-----setgpeasslltkvqgllts 88
XX
XX      70 ELIRKEIIEKIDOOIPPRP-----FTHVNTTTSATNST-----ATILNPDTCR 114
XX      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      89 ehkxwvdeevyk---ppeggsvfslrlrveathsgtqgtpesrlvnmctlsdd--cv 143
XX
XX      115 GDOLDILLEVRHDLGGRKQYGGDFLRA-----RMSSPALTAGASG-----KYMDFN 161
XX      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      144 ageidml-----gnglrlgrcvpyyqgpkctcevfqwcpevdaasvsgf-1 188
XX
XX      162 GTYLVSFTLFWEGVSLSLILHPSEGASALMRANOGYDK-----IIFK-GK 208
XX      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      189 gtmagnltlllkn--slhyprkfhskgnla--dtdgyllkrctfheasdlvcpilkgf 243
XX
XX      209 FVNGTSHVFTE-----CGTLNSNAELCEVLDRODEAFYC-----MKPOHMPCE 253
XX      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      244 lvekagesftelahnkgvlyglvlnwdcdl-----dlpasecmkpsftrldpkhvas 296
XX
XX      254 A 254
XX      297 s 297
XX
XX      RESULT 13
XX      AAAY70277
XX      10 AAAY70277 standard; Protein: 497 AA.
XX      AAAY70277;
XX      XX
XX      DT 20-JUN-2000 (first entry)
XX      XX

```

```

DE      Human purino receptor P2X2d (hP2X2dR) protein.
XX
XX      Purino receptor; human; hP2X2dR; ligand-gated ion channel; analgesic;
XX      auditory; agonist; antagonist; modulator; pain; treatment; vestibular;
XX      neuroendocrine disorder; gastrointestinal tract.
XX
XX      Homo sapiens.
XX      W0200011167-A2.
XX      02-MAR-2000.
XX      20-AUG-1999; 99MO-US19042.
XX      20-AUG-1998; 98US-0137458.
XX      (ABBO ) ABBOTT LAB.
XX      Lynch KJ, Burgard EC, Metzger RE, Niforatos W, Touma EB;
XX      WPI: 2000-237649/20.
XX      N-PSDB: AA251324.
XX
XX      Novel DNA encoding human P2X2 receptor, vectors and methods of
XX      identifying modulators of the receptor, useful for relieving pain and
XX      treating auditory, vestibular, neuroendocrine and gastrointestinal
XX      tract disorders.
XX      Claim 21: Fig 8D; 45pp; English.
XX
XX      The present amino acid sequence is the human purino receptor P2X2d
XX      (hP2X2dR) protein, derived from human pituitary gland tissues. P2X2
XX      receptors are ligand-gated ion channels, activated by ATP and other
XX      natural or synthetic nucleotides. It has analgesic and auditory activity.
XX      P2X2d sequences are useful for recombinant production and identification
XX      of therapeutic compounds, that modulate or interact with the P2X2
XX      receptor. P2X2d agonists and antagonists are useful for relieving pain
XX      and treating neuroendocrine, auditory and vestibular, and
XX      gastrointestinal tract disorders.
XX
XX      Sequence 497 AA:
XX
XX      Query Match 3.2%; Score 93; DB 21; Length 497;
XX      Best Local Similarity 22.9%; Pred. No. 1.4;
XX      Matches 69; Conservative 36; Mismatches 86; Indels 110; Gaps 18;
XX
XX      11 LLIISFSVVTMMIFISQNFKLSALNLSISVHYMNSAKSLFPKTSLI-PLKPLTET 69
XX      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      50 llllyl---vwyvllvqksyge-----setgpeasslltkvqgllts 88
XX
XX      70 ELIRKEIIEKIDOOIPPRP-----FTHVNTTTSATNST-----ATILNPDTCR 114
XX      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      89 ehkxwvdeevyk---ppeggsvfslrlrveathsgtqgtpesrlvnmctlsdd--cv 143
XX
XX      115 GDOLDILLEVRHDLGGRKQYGGDFLRA-----RMSSPALTAGASG-----KYMDFN 161
XX      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      144 ageidml-----gnglrlgrcvpyyqgpkctcevfqwcpevdaasvsgf-1 188
XX
XX      162 GTYLVSFTLFWEGVSLSLILHPSEGASALMRANOGYDK-----IIFK-GK 208
XX      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      189 gtmagnltlllkn--slhyprkfhskgnla--dtdgyllkrctfheasdlvcpilkgf 243
XX
XX      209 FVNGTSHVFTE-----CGTLNSNAELCEVLDRODEAFYC-----MKPOHMPCE 253
XX      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX      244 lvekagesftelahnkgvlyglvlnwdcdl-----dlpasecmkpsftrldpkhvas 296
XX
XX      254 A 254
XX      297 s 297
XX
XX      RESULT 14
XX

```


XX	Sequence	253 AA;
50		

Query Match	3.18;	Score 90;	DB 21;	Length 253;
Best Local Similarity	22.08;	Pred. No. 0.92;		
Matches	54;	Conservative	39;	Mismatches 103;
				Indels 50;
				Gaps 10;

```

07 14 LMSALNLSIVHWNNNSAKSLFPKTSLLPKKPLTELETRKEILEKIDDOI-PPPPFTHV 92
Db 15 LKSGSKSLTHLVLESKSL-----SLRLLSQTRFGENDTDLIDYDPPPYEQ 67
07 93 NNTTSATRSHTLINPDRTYCRGDGLDILEVDHLLGO-RROY-GPFLRANSSPALTA 150
Db 68 dlwcdlnrst-dlqeprr--fakrrrly-----kgrtkkmfsgwfl-hsnlckvkh1 117
07 151 GAGCWMPFNNNGYIVASFLLPFEROYSLSILLHPS-----EGASALMR 194
Db 118 llgltgltvbgqgtlsyrfhmsrvgqgnsvslyppkxlveidlaqgtvldakskstncr 177
07 195 ANNOGQIKLLFKKAFNGTSNHWFTBEGGLTNSNAELCEYLDDROEAFYCKMPQHPMPEA 254
Db 178 lcyeyvnda-----kntllcnypdskrcyqeqvltgshswlcsqpfvclcy 223
07 255 LTYMT 260
Db 224 lrlfyst 229

```

Search completed: October 12, 2001, 16:03:28
Job time: 51 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2001, 16:03:51 : Search time 20.85 seconds

(without alignments)
540.188 Million cell updates/sec

Title: US-09-729-454-2
Perfect score: 2908
Sequence: 1 MSSNTMLQKILLILISFSVY.....IHPDHYGNQIMFLNYIC 547

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: 1: Issued Patents AA: *
2: /cgn2_6/prodata/2/1aa/5A.COMB.pep: *
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/prodata/2/1aa/PCUS.COMB.pep: *
6: /cgn2_6/prodata/2/1aa/backfillseq1.pep: *

Fried. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	3.4	698	1 US-08-376-362A-20	Sequence 20, Appl
2	96.5	3.3	471	4 US-09-191-608-17	Sequence 17, Appl
3	94	3.2	404	4 US-09-191-608-18	Sequence 18, Appl
4	93	3.2	497	4 US-09-191-608-20	Sequence 20, Appl
5	90.5	3.1	582	4 US-08-194-560-2	Sequence 2, Appl
6	87	3.0	1164	2 US-08-589-756-1	Sequence 1, Appl
7	86.5	4.0	739	5 PCT-US93-00031-9	Sequence 9, Appl
8	86	3.0	1494	3 US-08-589-587-186	Sequence 186, Appl
9	85.5	2.9	447	4 US-09-191-608-19	Sequence 19, Appl
10	85.5	2.9	884	6 5208144-6	Sequence 4, Appl
11	85.5	2.9	3079	5 PCT-US94-00198-4	Sequence 2, Appl
12	85.5	2.9	3135	1 US-08-323-1708-2	Sequence 6, Appl
13	85	2.9	450	2 US-08-818-514-6	Sequence 6, Appl
14	85	2.9	450	4 US-09-115-934A-6	Sequence 6, Appl
15	85	2.9	766	2 US-08-553-436A-8	Sequence 8, Appl
16	85	2.9	2516	3 US-08-374-077C-2	Sequence 2, Appl
17	85	2.9	2516	4 US-08-895-590-2	Sequence 2, Appl
18	84.5	2.9	1150	2 US-08-589-756-3	Sequence 3, Appl
19	84.5	2.9	3418	3 US-08-639-501-2	Sequence 2, Appl
20	84.5	2.9	3418	3 US-09-044-946-2	Sequence 2, Appl
21	84.5	2.9	3418	3 US-09-044-908-2	Sequence 2, Appl
22	83.5	2.9	318	3 US-09-188-579-81	Sequence 81, Appl
23	83.5	2.9	318	4 US-09-315-444-81	Sequence 81, Appl
24	83.5	2.9	324	2 US-08-528-057-46	Sequence 46, Appl
25	83.5	2.9	1152	2 US-08-476-062A-43	Sequence 43, Appl
26	83.5	2.9	1152	5 PCT-US96-01314-43	Sequence 43, Appl
27	83.5	2.9	1152	6 5424399-2	Patent No. 5424399

28	83.5	2.9	1153	1 US-08-173-497-3	Sequence 3, Appl
29	83.5	2.9	1153	1 US-08-286-889-3	Sequence 3, Appl
30	83.5	2.9	1153	1 US-08-485-618-3	Sequence 3, Appl
31	83.5	2.9	1153	1 US-08-362-652-3	Sequence 3, Appl
32	83.5	2.9	1153	2 US-08-605-672-3	Sequence 3, Appl
33	83.5	2.9	1153	2 US-08-482-293A-3	Sequence 3, Appl
34	83.5	2.9	1153	2 US-08-943-363-3	Sequence 3, Appl
35	83.5	2.9	1153	4 US-09-193-043-3	Sequence 3, Appl
36	82.5	2.8	1050	3 US-09-045-632-49	Sequence 49, Appl
37	82.5	2.8	1050	3 US-08-589-756-2	Sequence 50, Appl
38	82.5	2.8	1167	2 US-08-677-734A-9	Sequence 9, Appl
39	82	2.8	834	2 US-08-677-734A-10	Sequence 10, Appl
40	82	2.8	2329	3 US-08-755-587-16	Sequence 16, Appl
41	82	2.7	1075	1 US-07-623-033-2	Sequence 2, Appl
42	79.5	2.7	3418	2 US-08-603-753D-4	Sequence 4, Appl
43	79.5	2.7	3418	3 US-08-755-587-44	Sequence 44, Appl
44	79.5	2.7	3418	4 US-09-099-753-4	Sequence 4, Appl
45	79.5	2.7	3418	4 US-09-099-753-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-376-362A-20
Sequence 20, Application US/08376362A
Patent No. 5693756
GENERAL INFORMATION:
APPLICANT: Li, Xiao-Jiang
APPLICANT: Snyder, Solomon H.
TITLE OF INVENTION: AMILORIDE-SENSITIVE SODIUM CHANNEL AND
TITLE OF INVENTION: METHOD OF IDENTIFYING SUBSTANCES WHICH STIMULATE OR BLOCK
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, N.W., Eleventh Floor
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20001-4597
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/376,362A
FILING DATE: 23-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan A., Sarah
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,48125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 698 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rattus rattus
US-08-376-362A-20

Query Match 3.4%; Score 98; DB 1; Length 698;
Best Local Similarity 19.2%; Pred. No. 0.28;

Matches 89; Conservative 53; Mismatches 157; Indels 164; Gaps 18;

QY 11 LLLISFSVVTWMIIFISQNFTKLMSALNLSISVHYNNNSAKSLPRTSLPLKPLTERE 70
 DB 114 VLMLCTFGMYWQFALLFEELSYSPVSLINL-----NSDKLVFPAVTCILNPRTE 167
 QY 71 LRIKEIEKLDQ-----QIPPRPF 89
 DB 168 ---IKEELEEDRLTEOTLDLYKKNSSYTOAGARRSSRDLLAFPHPLQLRLRPPPPY 225
 QY 90 THVNTTTSATHSTATLNR-----DTYCGDDLDLLEVDH 127
 DB 226 S---GRTASGSSSVRDNNPOYDRKDMKIGPOLCNKSDCFYQTSYG--VDAYRE---- 277
 QY 128 LGQRKQYGDPLFARMSPALTAGSGKVDENNQYLVSTLFMEQVSLSLILHSE 187
 DB 278 -MYRHYINILSLSDTSPALEEALG-----NFLTCFRNQAQMAVYS-KFHHHPY 329
 QY 188 G-----ASALPARNQGYDKIIFKGFVNGTSHVTEGGLTNSNAELCEYLDPD 238
 DB 330 GNCYTFENDKNNSLMWSMPG-----VNN-----GLSLTLTEQNDFI---- 367
 QY 239 QEAFCMKPQHMPCALTYMTTRNREVSYLTDKENSFLHRSKVGVEMMDRKHIDVTNCN 298
 DB 368 -----PLSLVTGARVWVGDDFAFDGCGNLPRVETISKKREA--LDSLGCN 417
 QY 299 KREKIEETCOYGMKPPVPGGTTLOGKMTTFPCNOYOLDTIRKNGCKGLIYLDSTLR 358
 DB 418 YGDCYENGSDVPYKMLYPSKYTQ-----VCIHSCQDEMMIKKCGCA--- 459
 QY 359 QMIYY-FPKVVKTLKF-----FDLHETGIFPK 384
 DB 460 -YIFYPKKGVCEFCOYBRKQSSMGVYCYKLGAFSLDLCGCSK 501

RESULT 2
 US-09-191-608-17
 ? Sequence 17, Application US/09191608
 ? Patent No. 6242216
 ? GENERAL INFORMATION:
 ? APPLICANT: Lynch, Kevin J.
 ? APPLICANT: Burgard, Edward C.
 ? APPLICANT: Metzger, Randy E.
 ? APPLICANT: Nitrosatos, Wende
 ? APPLICANT: Touma, Edward B.
 ? APPLICANT: Van Biesen, T.
 ? TITLE OF INVENTION: Nucleic Acids Encoding a Functional
 ? TITLE OF INVENTION: Human Purinoceptor P2X2 and P2X4 And Methods Of Production
 ? FILE REFERENCE: 6394, US, P1
 ? CURRENT APPLICATION NUMBER: US/09/191,608
 ? CURRENT FILING DATE: 1998-11-13
 ? NUMBER OF SEQ ID NOS: 26
 ? SOFTWARE: FastSeq for Windows Version 3.0
 ? SEQ ID NO 17
 ? LENGTH: 471
 ? TYPE: PRT
 ? ORGANISM: Homo sapiens
 US-09-191-608-17

Query Match 3.38; Score 96.5; DB 4; Length 471;
 Best Local Similarity 21.4%; Pred. No. 0.2;
 Matches 86; Conservative 46; Mismatches 120; Indels 157; Gaps 23;

QY 11 LLLISFSVVTWMIIFISQNFTKLMSALNLSISVHYNNNSAKSLPRTSLPLKPLTERE 69
 DB 50 LLLILYF---WYVFIYOKSYD-----STGPESSITTKKAGTTIS 88
 QY 70 ELRIKEIEKLDQIIPRP-----FTHVNTTTSATHST-----ATLNRDVTYR 114
 DB 89 EHKVMDVEEYK---PPEGGSVFSIITRVEATHSQGTGCPESIRVHNATCLSDAD--CV 143

QY 115 GEODLLEVRDHILGQRKQYGDPLFAR-----RMSSPALTAGAGS-----KYMDFRN 164
 DB 144 AGELDML-----GNGLRIGRCVPIYQGPSTCEVGMCPEDGASVSLF-L 186
 QY 162 GTYLVSTLFMEQVSLSLILHPSGASALMARNOGYK-----IIFK-GK 208
 DB 189 GTMAPNFTILIKN--SIHPKFRHSKANI-----DRTDGYLKCTFHEASOLYPIPIKLCG 244
 QY 209 FVNGTSHVTE-----CGTLNSNAELCEYLDROGEAFYC-----KKVLMHCE 253
 DB 244 IVEKAGESEFTELAKHGQVIGYIIMMODL-----DLPASECNKISFRLLDPKHPVAS 296
 QY 254 A-----LTYMTTRN-----REVSYLTDKENSFLHRSKVYV-E 294
 DB 297 SGYFRFPAKYKINGTTRTLIKAYGIRIDVIHQAGKFSLLPTIINLATALTSVGVS 356
 QY 285 MKDKRHIDVTNCN---REKIEETCO-----VGMKPPYNG 317
 DB 357 FLCDDMLTFPMKNNKKNVYSHKKFEDKVCPTPSHSGSPVTIARVLSLAPERRG 407

RESULT 3
 US-09-191-608-18
 ? Sequence 18, Application US/09191608
 ? Patent No. 6242216
 ? GENERAL INFORMATION:
 ? APPLICANT: Lynch, Kevin J.
 ? APPLICANT: Burgard, Edward C.
 ? APPLICANT: Metzger, Randy E.
 ? APPLICANT: Nitrosatos, Wende
 ? APPLICANT: Touma, Edward B.
 ? APPLICANT: Van Biesen, T.
 ? TITLE OF INVENTION: Nucleic Acids Encoding a Functional
 ? TITLE OF INVENTION: Human Purinoceptor P2X2 and P2X4 And Methods Of Production
 ? FILE REFERENCE: 6394, US, P1
 ? CURRENT APPLICATION NUMBER: US/09/191,608
 ? CURRENT FILING DATE: 1998-11-13
 ? NUMBER OF SEQ ID NOS: 26
 ? SOFTWARE: FastSeq for Windows Version 3.0
 ? SEQ ID NO 18
 ? LENGTH: 404
 ? TYPE: PRT
 ? ORGANISM: Homo sapiens
 US-09-191-608-18

Query Match 3.28; Score 93; DB 4; Length 404;
 Best Local Similarity 22.9%; Pred. No. 0.37;
 Matches 69; Conservative 36; Mismatches 86; Indels 113; Gaps 14

QY 11 LLLISFSVVTWMIIFISQNFTKLMSALNLSISVHYNNNSAKSLPRTSLPLKPLTERE 69
 DB 50 LLLILYF---WYVFIYOKSYD-----STGPESSITTKKAGTTIS 88
 QY 70 ELRIKEIEKLDQIIPRP-----FTHVNTTTSATHST-----ATLNRDVTYR 114
 DB 89 EHKVMDVEEYK---PPEGGSVFSIITRVEATHSQGTGCPESIRVHNATCLSDAD--CV 143
 QY 115 GDDDLLEVRDHILGQRKQYGDPLFAR-----RMSSPALTAGAGS-----KYMDFRN 164
 DB 144 AGELDML-----GNGLRIGRCVPIYQGPSTCEVGMCPEDGASVSLF-L 186
 QY 162 GTYLVSTLFMEQVSLSLILHPSGASALMARNOGYK-----IIFK-GK 208
 DB 189 GTMAPNFTILIKN--SIHPKFRHSKANI-----DRTDGYLKCTFHEASOLYPIPIKLCG 244
 QY 209 FVNGTSHVTE-----CGTLNSNAELCEYLDROGEAFYC-----KKVLMHCE 253
 DB 244 IVEKAGESEFTELAKHGQVIGYIIMMODL-----DLPASECNKISFRLLDPKHPVAS 296
 QY 254 A 254

Db 297 S 297

```

RESULT 4
US-09-191-608-20
: Sequence 20, Application US/09191608
: Patent No. 6242216
: GENERAL INFORMATION:
: APPLICANT: Lynch, Kevin J.
: APPLICANT: Burgard, Edward C.
: APPLICANT: Metzger, Randy E.
: APPLICANT: Niforatos, Wende
: APPLICANT: Touma, Edward B.
: APPLICANT: Van Biesen, T.
: TITLE OF INVENTION: Nucleic Acids Encoding a Functional
: TITLE OF INVENTION: Human Purinoreceptor P2X2 and P2X4 And Methods of Production
: FILE REFERENCE: 6194 US P1
: CURRENT APPLICATION NUMBER: US/09/191,608
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 20
: LENGTH: 497
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-191-608-20

```

Query Match 3.2%; Score 93; DB 4; Length 497;
 Best Local Similarity 22.9%; Pred. No. 0.52;
 Matches 69; Conservative 36; Mismatches 86; Indels 110; Gaps 18;

```

QY 11 LILISFVWMMIFLISQFTKMSALNISIVHYNNNSAKSLPKPLLET 69
Db 50 LILILF---VWVFLVQKSYOE-----SETGPESSITTKYKGLTTS 88
QY 70 ELRIKEIIEKLDQOIPRP-----FTVNTTTSATST-----ATLINPROTYCR 114
Db 89 EHKVMDVEEYK---PREGSVFSITTRVATHSQTGCTPESIVHNAITLSAD--CV 143
QY 115 GQDLILLEVDHIGRKQKQYGDPLRA-----RMSFALTAGASG-----KWDFFNN 161
Db 144 AGELDML-----GNGLRTGRCVPIYQGSKICEVFGMCPVEDGASVSQF-L 188
QY 162 GTYLVSTFLFWEGCVSLTLHPSEGASALMARANQGYDK-----ITFK-GK 208
Db 189 GTMAPNFILIKN--SIHYKFFHFKGNIA---DRTDGYLKCTFHEASDLYCPITFLGP 243
QY 209 EVNGTSHVETE-----CGLTNSNAELCEYLDHDOEAFYC-----MKQGHMPC 253
Db 244 IYKKGESFTFLAHKGVGIVGIIMDCDL-----DLPASECNKYSFRRLDPKHPAS 296
QY 254 A 254
Db 297 S 297

```

```

RESULT 5
US-08-194-560-2
: Sequence 2, Application US/08194560
: Patent No. 6255062
: GENERAL INFORMATION:
: APPLICANT: Campbell, Judith L.
: APPLICANT: Budd, Martin E.
: TITLE OF INVENTION: B-Type DNA Polymerases
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fiehn, Hobbach, Test, Albritton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States

```

```

ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/194,560
: FILING DATE: 14-FEB-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Trecartin, Richard F.
: REGISTRATION NUMBER: 31,801
: REFERENCE/DOCKET NUMBER: A-59515/RFT/RMS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 582 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-194-560-2

```

Query Match 3.1%; Score 90.5; DB 4; Length 582;
 Best Local Similarity 20.6%; Pred. No. 1.3;
 Matches 52; Conservative 40; Mismatches 80; Indels 81; Gaps 12;

```

QY 197 NCQYDK-----IIRKGFVNGTSHV-----FTGCLTNSNAELCEYLD 236
Db 359 NRGYSKGDIDLFFKPCNDTTELAKIMETLICKIKYKGYHICFLQITPYLE----- 411
QY 237 RDEAFVYKMPQHPCEALTYMTNRNRSVLTQKENSJFHSKY-----GV----- 283
Db 412 ---KLF-----LKRIVERFRTAKIVGGERKKWYSSELIKFFPKYKFSPREL 456
QY 284 EMKDKRKHIDVTNCKREKIEETCOVGKMPVPGVYTLQKWIITFCNOVOLDTIRINGC 343
Db 457 EELKEKKNDDEGTLIDEEDEEET---KLNPIDQYSLNAK-DGNYCRHLDFFCCWDEL 511
QY 344 LKGLIYLLGDSLTRQWYIFPVVYKTLKFDLHETGIFKKLLDA--ER----- 392
Db 512 GAGRHY-TGSEKYNRWI-----RIIAQKGYKLQHGDLFRNNILLESINERIFELLNK 566
QY 393 -----HTQIQKK 400
Db 567 YAEPEHRNIEWEK 579

```

```

RESULT 6
US-08-589-756-1
: Sequence 1, Application US/08589756
: Patent No. 5846547
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: STREPTOCOCCAL C5a PEPTIDASE VACCINE
: NUMBER OF SEQUENCES: 5
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/589,756
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```

MOLECULE TYPE: protein
US-08-589-756-1

Query Match

Best Local Similarity 3.0%; Score 87; DB 2; Length 1164;

Matches 93; Conservative 49; Mismatches 132; Indels 168; Gaps 24;

```

QY 144 SSPALTAGASGKYMDENNGTYLVSFT---LFMEGOVSLILLHPSEGASALMRANQY 200
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 450 SQTITFNATPKVLPYASTGKLSRFSMSGTLADGNKPD---MAGCDILSSANKY 505
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 201 DKLIIFKGFVNGTSH---VFTEGCL-----TLNSAEICE-----YLLD 236
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 506 AKL-----SGTSMAPLAVAVINGLLQKQYETQYDPTQTOSERLDLAKVLMSSATALY 558
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 237 RDOEAFYCKMPQHMPCALTYMTNRNREVSYLTDKENSILPHRSVGVEMMKRKHIDTN 296
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 559 EDEKAYE--SPRGQAGAVD-AKKASATMYTQKDN---SKVHLNNVSDKEVETVY 612
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 297 CNKREKIEE---TCQYGMKPPVPGY-----TLQGM--ITFCNQVQDLTIKING 342
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 613 HNKSDKPHELYQATVOT---DKYDKHFLAPKALLETSMOKITTPANSSKQVTIPI-- 667
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 343 CLKGLIYLLGDSTLRQWYVFPVYVTKLPFLDHEGTGIFKKHLLDAR-----RH 393
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 668 -----DISQ---FSKOLLQMKNGYFLEGVRI 692
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 394 TOIQMKH--SYRPTVQ-----LXSLID---HDYIREIDRLSD----- 429
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 693 KQPTKEELMSIPYIGFGDQGNLSALENRPYDSKSSYHEEISDAKDQDLDDGLQFY 752
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 430 --KNTAVITFGQHRPPIDIFIRRAIGVQKALERFLRSPATKVIKTEIREMH--- 484
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 753 ALKNDFTALT--TESNPTIINVKEG-----ENIDDISSSE 788
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 485 -IETEEFGDF-----HGYIH 498
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 789 ITETIFAGTFAKODDRHYIHH 810

```

RESULT 7

PCT-US93-00031-9

Sequence 9, Application PC/TUS9300031

GENERAL INFORMATION:

APPLICANT: Osborn, Laurelee

APPLICANT: Benjamin, Christopher D.

TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: Allegretti & Wilcoff, Ltd.

STREET: 10 South Wacker Drive

CITY: Chicago

STATE: IL

COUNTRY: US

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/00031

FILING DATE: 19930112

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: McNicholas, Janet M.

REGISTRATION NUMBER: 32,918

REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 715-1300

TELEFAX: (312) 715-1234

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 739 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US93-00031-9

Query Match

Best Local Similarity 3.0%; Score 46.5; DB 5; Length 739;

Matches 88; Conservative 73; Mismatches 170; Indels 119; Gaps 21;

```

QY 11 LLLIISTVYTMIFITISQNF-----TKMSALNLSIVHYNNNSAKSLF----- 74
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 5 MVLILASNLIMLIMFMAASQAFKIETTPESRYLAQIGDSVSLTISTESTFETKRGDID 44
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 56 -----PKSLIPLKPL-----TELRLKELIEKLAVIHPHETRYN 94
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 65 SPLNGKVTNEGTTSTLIMNPFSGNHSYLCATCESKLEKGLQVEYVFYKDFEHLHS 124
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 94 TTTSATSHATILANPRTGYCRGDQDILLEVVDHLGQRQYGNFLAAMSSALTAGAS 164
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 125 GPLEACKPIYKQSVADYV--PFDRLIEDLLKGDHLMKSQFLEDAKRSLETKSLEVTY 164
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 154 GKVMDFNNGTYLVSFTLFW-----EGGVSL-----LLIHS-----EAS 164
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 184 PVIEDI--GKVLVCRAKLHIDEMDSVPTVQVAVKELOVYISPRNTVIVSNSTYKLGSS 241
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 191 ALMRANQY--DKLIIFKGFVNGT--SHVTEGCLTL-----NSNAELTEYL----- 246
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 242 VYTKTSSEGLPAPLFIKSKLQNGNLQHLQSNATLTILAMRMEDSGIVVCGVNLQKRN 401
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 237 RDOEAFYCKMP-----QHPCALTYMTNRNREVSYLTIK-- 271
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 302 KEVELIVQEKFPFVEISPGPIAAQIGDSVNLGCSVNGCISPSF--SMRIQVDSLSKYN 466
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 272 ---ENSLFHRSKVGEVEMKDRKHIDVTGCKNREKIEETCVGM-----KRPVP-GVYTLQ 422
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 361 SEGTNSTLTLSPVSE--NEHSYLCVTVCG--HKRLKGLQVELYSFPRDEIEMSGLVN 417
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 323 GKMITFC---NOVQDLTIKINGCLKGLI 349
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 418 GSAVIVSCKPSVYPIDRLIE-LKGETI 446

```

RESULT 8

US-08-755-587-186

Sequence 186, Application US/08755587

Patent No. 6045997

GENERAL INFORMATION:

APPLICANT: Futreal, Phillip A

APPLICANT: Wooster, Richard F

APPLICANT: Ashworth, Alan

APPLICANT: Stratton, Michael R

TITLE OF INVENTION: Materials and methods relating to the

IDENTIFICATION AND SEQUENCING OF THE BRAIN GENE

TITLE OF INVENTION: Identification and sequencing of the BRAIN GENE

NUMBER OF SEQUENCES: 222

CORRESPONDENCE ADDRESS:

ADDRESSEE: Bell Seltzer Park & Gilson

STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107

CITY: Raleigh

STATE: NC

COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPD)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/755,587

FILING DATE: 25-NOV-1996

```

1 PRIORITY APPLICATION DATA:
2 PRIORITY APPLICATION NUMBER: GB 9523959.6
3 FILING DATE: 23-NOV-1995
4 PRIORITY APPLICATION DATA:
5 PRIORITY APPLICATION NUMBER: GB 9525555.0
6 FILING DATE: 14-DEC-1995
7 PRIORITY APPLICATION DATA:
8 PRIORITY APPLICATION NUMBER: GB 9617961.9
9 FILING DATE: 28-AUG-1996
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Kenneth D. Sibley
12 REGISTRATION NUMBER: 31,665
13 REFERENCE/PRIORITY NUMBER: 5,405,135
14 INFORMATION FOR SEQ. ID NO. 186:
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 1494 amino acids
17 TYPE: amino acid
18 TOPOLOGY: linear
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
10
```

Query Match 3.08; Score 86; DB 3; Length 1494;

Best Local Similarity 22.0%; Pred. No. 19;
Matches 65; Conservative 45; Mismatches 128; Indels 58; Gaps 14;

```

QY 35 MSAALULSTSVHYWNWNNKSKLSPFKTSLIPKLPLETETLRJKEIETIEKLDQIPPPETHYNT 94
Db 354 MARELDPIINSHFENFGCTAKSKKEIKLENNIKRSMKLEIDIE-----HYPT 400
QY 95 TISA-----TSTATILINRDYTCGCGDOLITVBDHGQKQYCGDPLARMSSEPAL 14
Db 402 NIACLEIYNTSSLEQKKRKSCHALDPINII-----SGVQNSTYVDSESGHTAPPLIS 45
QY 150 AACASQVWDFNKGTYLVSTLFWQGVQY--LSLLIHPSGCAALMRANQCYKIIFFKR 20
Db 458 L-----KQDFDSRNKL-----TPSQAKETIELSTIL-----ESSGQPEFTQFKRPSHIIQKAP 50
QY 209 F-----VNGTSHVTFECGLTINSAELCEIYLDROQEAFTCMKPDQHMPEALTYTT 25
Db 507 FEMPEQNTLITNLSCKEMKDDLLHTTAPSIQ-VQSKSSEGIIGCK-QKFACLSS---RT 56
QY 260 TNNREVS-VYTDKENSLFH-----NSKVGYEMMKDKRHH--DVYCNKREKIE 304
Db 562 SCNRASQSYSDTKKNEVEFRGFYSARGTGLNVSALALQKKKLFQSDLENIIEETSV 617

```

RESULT 9
US-09-191-608-19

```

Patient No. 6242216
GENERAL INFORMATION:
APPLICANT: Lynch, Kevin J.
APPLICANT: Burgard, Edward C.
APPLICANT: Metzger, Randy E.
APPLICANT: Niforatos, Wende
APPLICANT: Touma, Edward B.
APPLICANT: Van Biesen, T.
TITLE OF INVENTION: Nucleic Acids Encoding a Functional
TITLE OF INVENTION: Human Putinoceptor P2X2 and P2X4 And Methods Of Production
TITLE OF INVENTION: And Use Thereof
FILE REFERENCE: 6394.US.P1
CURRENT APPLICATION NUMBER: US/09/191,608
CURRENT FILING DATE: 1998-11-13
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PASTSEQ for Windows Version 3.0
SEQ ID NO 19
LENGTH: 447
TYPE: prt
ORGANISM: Homo sapiens
US-09-191,608-19

```

Query Match 2.98; Score 85.5; DB 4; Length 447;

Best Local Similarity 20.7%; Pred. No. 2.7;
Matches 82; Conservative 46; Mismatches 117; Indels 151; Gaps 21;

Matches	82;	Conservative	46;	Mismatches	117;	Indels	151;	Gaps	21;
---------	-----	--------------	-----	------------	------	--------	------	------	-----

11 LLLISFSVTTMFIISQNTKLMASLNLISVHYNNNSAKSLFPKTSLI-PLKPLTER 69

```
b      | : | : | : | : | : | : |  
50 LLILLYF--VVVFIYKSIQE-----SEIGPSSIIYVKGITTS 88
```

```

70 ELRIKELI EKDOIPRPRTHTNTTSATHTATILNPRDTYCRGDOLDILLEVRDHLG 129
  | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |

```

89 EHKVDEE-----YKPPESIRVHNATCSDAD--CVAGELDML----- 126

```

130 QKQIGDFLRA-----RMSPALTAGAG-----KVMDFNNGTYLVSTLFMEGV 176
      | : | | : | | : | | : | | :

```

127 -----GNGLRTRGCAVPYVQGSKTCFVGMCPVEDCASVSQF-LGTMAPNFTILIKN-- 177
177 CTTCTTTDDEGCAATMAPNCGVDK-----TTRK-CRWNGCTSHVETE-... 310

[illegible]

220 ---CGTTLNSNAELCEYLDDRDQEAFFYC-----MKPQHMPCEA-----LTY 257
178 SIRTIRFRFSAGNIA---DRIDGILNRCIFREASDLCLIFKNDQFVENWDSFIEEMNN 234

235 GGIVGVIINWDCDL-----DLPASECNPKYSFRRLDPKHPVAPSSGYNFRFAKYYKING 287

258 MTTRN-----REVSYLTDKENSLEFRSKGV-EMMKDRKHIDVTNCNK 299

288 TTRRLIKAYGIRDIVHGGAGKESLIPTIINLATATLSVGVGSPICDMLITFMNKK 347

300 ---REK1ETCQ-----VGMKPPVPG 317

348 VYSHKKFDKVCPTSPHSGSWPYTLARVLGQAPPEPG 383

RESULT 10

208144-8
Patent No. 5208144

APPLICANT: SMITH, JOHN A.; RAYCHOWDHURY, RAJTIMA; NILES, JOHN L.
TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA
CONTAINING THE GENE ENCODING FOR DENSITY ITROPOTEIN RECEPTOR

4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 5

CONTRACT AFFILIATION DATA:
APPLICATION NUMBER: US/07/396,697
FILING DATE: 22-AUG-1989

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 313,682

FILING DATE: 22-FEB-1989
 APPLICATION NUMBER: 235,211

FILING DATE: 23-AUG-1988
SEQ ID NO: 8
INVENTOR: 004

LENGTH: 884
5208144-8

Query Match 2.98; Score 85.5; DB 6; Length 884;

Best Local Similarity 22.0%; Pred. No. 8.6;
Matches 41; Conservative 17; Mismatches 69; Indels 59; Gaps 10

159 FNNGYL-----VSFTLFW-EGQ-----VSLSLLIHPSEGASALMRARNQ 198

87 FPNGTGCPAMFECKNHVCIOSEWICDGENDCVDSDEIHLFCNIPCESPQR-FRCDN-144

199 GYDKIIFRGKGVNGTSHVFTECGLTLSNAELCEYLDLRDQDAFYCKMPQHPCEALTYM 258

145 --SRCVYGHQLCNGVD---DCG-----DSGDEKEHCKKPTHKPCPTDEYK 185

259 TTRNREVS-YLTDKENSLFHRKSGVGEEMKDKRKHIDVYNCN---KREKIEETQVGMKP 313

186 CSNGNCISQHYVCNVDGGLS-----DEIGCNLGDNTCAENICEQNCTQ 232

2Y 314 FVPGGX 319
||:
222 TCCCCC 228

233 LSSGGF 238

RESULT 11
PCT-US94-00198-4
Sequence 4, Application PC/TUS9400198
GENERAL INFORMATION:
APPLICANT: Schering Corp.
TITLE OF INVENTION: RAS Associated GAP Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering Corp.
STREET: 1 Giraldo Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 6.0.8
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00198
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/004,824
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lunn, Paul G.
REGISTRATION NUMBER: 32,743
REFERENCE/DOCKET NUMBER: DX0352 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)822-7255
TELEFAX: (201)822-7039
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3079 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
PCT-US94-00198-4

Query Match 2.98; Score 85.5; DB 5; Length 3079;
Best Local Similarity 19.3%; Pred. No. 72;
Matches 93; Conservative 85; Mismatches 160; Indels 143; Gaps 27;
QY 3 SNTLQKTLILISFSVYTWMIIFISQNTKLMASLNL--SISVHYMNSAKSLFPKTSL 60
DB 2515 ANNIOKILAKIRSF-----LPSLKEAMTQSMSELILVAKISIH-----VFETSL 2561
QY 61 IPLKPLTELRLKIEIKEDQIIPRPPTVNTTSTATHTIIN--PROTYCGDQ 117
DB 2562 LVQKYLPELFIIVSLIDVGPRL--RSSLLH-QLIMNVCHSLA--INSALPD--HRN 2613
QY 118 LDILLEVRDHLQORQYGGDFLARNSS--PALTAGASGVMDF--NGQYIVSFTLFE 173
DB 2614 LDEISDIFAHQVKFMFESFEDGRILQIFSSASKFNILDFINILLIMESSTYE 2673
QY 174 GOVSLSLILHPSSEGASALWRANQGYDIIFKCKGFNGSHVFTECGTLTNSNAELCE 213
DB 2674 ANV-----WKTR--YKKYVLESVF-----TNS-----F 2695
QY 234 LDDRODAFYCKKPOHNP---CEALTYYMTRNREVSYLTKK-----NSLFHRSKV--- 281
DB 2696 LSAKSIIMVIGIMKSYITEGLKAMLIETMVAIEPKITDEHLFLAISHIFTYSKIVEGL 2755
QY 282 --GVEMKK-----DRKH-----IDVTNCKNR-----EKLEECQGVKMPV 315
DB 2756 DPNLDLMKHLFWFTLFLSRHPIIPEGALLFVNCICRLRYMAQFENSESTLSLT----- 2811

QY 316 PGSVYLLG-KWITTFONOVQDITKINGCLGK-----LAVLSDILKQWYFFKRV 408
DB 2812 -----LKGRFAHFTLSKIE-----NLGIWNMEDVFTIILINGLSNRP1----- 2674
QY 369 KILKFDLHETGIFKKHLLDAEHRQIQMKKHSIPVTTQVLSLDHYVIFHEIKSS 424
DB 2857 KSTAF-----DFLKMFRNSYFEHQIQKSDHYLCVMPLYVANCN-----DFEELG 2906
QY 429 D 429
DB 2906 D 2906
RESULT 12
US-08-323-170B-2
Sequence 2, Application US/08323170B
Patent No. 5733772
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
TITLE OF INVENTION: Cloning and Expression of Plasmidum
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-1131000US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 315 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-323-170B-2

Query Match 2.98; Score 85.5; DB 1; Length 315;
Best Local Similarity 19.9%; Pred. No. 74;
Matches 106; Conservative 75; Mismatches 168; Indels 185; Gaps 26;
QY 27 ISQNTKLMASLNLISVHYMNSAKSLFPKTSLIPKPLTEFLRIKIEIKEDQIIP 46
DB 943 ICRKHLKEDFTIKCNVNTQVNPNIETFPPT-----LKKRKEVLKIDLDIY 990
QY 87 RPT---HVNITTSATSTATILNPRDYCGDGLDILLEVRHLDQKQYGGD----- 137
DB 991 QMSKFFKFNTP-----QNAKYLNTXPY-----LFFPNHIGKMKLKNFTYKNR 1036
QY 138 -----FLRKMSSPALTAGASGVMDF-----SNITTVVSF 164
DB 1037 DVKFEQSSVLSPLSSADSLGKILNLDIOTVCLTIKIRYLNLSINELSDNNTSVTF 1096

```

QY 169 TL-----FW-----EGVSLSLLLHPSEGASALMRARNOGDKITFGKFT 209
DB 1097 QVHPYIDIKEPFYFPGCNKNKGDNIGIVELLISKQEKFT--GCGFHSKSKLDYNNEN 1153
QY 210 VNQTSHTFEGCLTLNSNAELCEYLDRODAFYCMKPOHMPCE-----ALTYMTTN 262
DB 1154 ISSDTH--ECTLHAYEN-----DIIGFNGLETTH-PREVEVEVDAELIYLOPEN 1199
QY 263 -----REVSYLTDKENSLEFHSKSGVEMMKDRKHIDYTNCKRREKIEETQOVAMK 313
DB 1200 CENNYVKGLNSVDITTLKNA-----QTYNNKKTPTFLKIPYINLEDEVELSCQTIKO 1255
QY 314 PYRGYTLQGMWITTFPCNOYQDITKINGCLKGLIYLDGDSLROMIY-----YFPK 366
DB 1256 VVW-----KIKVILIT-----KNDTV-----VQSESTLDDKIKCEHENFINPR 1298
QY 367 VVWTL-----KFFD-----LHETGIEFK-HLLDAERHTQI-QMKKHSYF 404
DB 1299 VNKTFDENVEYTCNIKIEFNFTIOIFCPAMDQIYNNIOMIYDIYKPRVPOFKK---- 1354
QY 405 FYTFQLYSLIDHDYIPREDIRLSGDKNTAIVITFGQ-----HFRFPDIFIR 452
DB 1455 FNNELHMLIPNSEMLH-----KITKEMILYNEEKVULLHFYVVF-LPIYIK 1399

RESULT 13
US-08-818-514-6
; Sequence 6, Application US/08018514
; Patent No. 5817838
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: BAX Inhibitor Proteins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,514
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2446
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 450 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-818-514-6

Query Match 2.9%; Score 85; DB 2; Length 450;
Best Local Similarity 18.9%; Pred. No. 3.1;
Matches 76; Conservative 54; Mismatches 133; Indels 140; Gaps 18;
DB 46 YNNNSKSL-----FPKTSLPLKPLTELKIKETIKLDDQIIPRPTHTYNTT 95
DB 61 WNASKRIEPCREKKEWGFVKVSI-----LRDAIEKLFPALIRLFREDIQN 109

```

```

QY 96 TSATHTSTATILNPDYTCRGDQDILLFVRDHLGQ-RKQYGDGLFARMSPALTAGASG 154
DB 110 NDVQSLAFOK-----YGNQDIPL---APNIGRANQOMGGGFF-----SGVLT----- 151
QY 155 KWMPDFNGCTVYVSFTFLFMEGOVYSLLLIHPSEGASALMRARNOGYDKIIFKCPFNCTS 214
DB 152 -----LTGVAVYVLLYHMSRSRESHDLVHK---AAAKTA-----EEVYIMLEQGPRA 198
QY 215 HYTEGCLTLNSNAELCEYLDRODAFYCMKPOHMPCEALTYMTTNREVSYLTDKENS 274
DB 199 SLRE-----RFLSR-----VNCRLILLTLEF-- 222
QY 275 LFHSKSGVEMMKDRKHIDYTNCKRREKIEETQOVGKPPVGYTLQGMWITTFPCNOYO 334
DB 223 -FCKPTYLTIENSSHRAILM-----ELEVKALGVKPP-----QNLN----- 258
QY 335 LDTIKNGCLGKLIYLDGDSLROMIYFPKVKYTKLKFQDLHETGIFKKHL----- 386
DB 259 -EYKAVNPGRSLLFLYALKSSPRLSLY-----LYLEDYDTDFLFPHTICPQEDS 309
QY 387 -----LLDAERHTQIQMKH--SYFVTFQLYSLIDHDYI 419
DB 310 SGEDIYTKLDDKEPTWKQWERFLVYGFLLYQULIAEFAMDWL 352

RESULT 14
US-09-115-934A-6
; Sequence 6, Application US/09115934A
; Patent No. 6130317
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; TITLE OF INVENTION: BAX Inhibitor Proteins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,934A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/818,514
; FILING DATE: 14-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3209
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 450 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-09-115-934A-6

Query Match 2.9%; Score 85; DB 4; Length 450;
Best Local Similarity 18.9%; Pred. No. 3.1;
Matches 76; Conservative 54; Mismatches 133; Indels 140; Gaps 18;

```

[illegible]

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-553-436A-8

```

Query Match 2.98; Score 85; DB 2; length 766;
Best Local Similarity 19.98; Pred. No. 7.6;
Matches 62; Conservative 56; Mismatches 131; Indels

```

QY 264 EYUJTEKSNSTPHSKGVEEMK---DKNHLD-----INCKR-EXT- 314
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 QIVYILHDVNSLH-----EMDHIKOGGLDTPRILVSELIWAKGTJNMEKVS 300
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 305 ETCQVG-MKPPVGGYTLQGMKWTTFPNQVOLDTI-----KINGTJCKEJLYLGEKTL 375
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 301 GTEHNSILRPVFESEKGLIRKWSIFPMWRYETLEFEDAGEJLIGELGSPOLLGN--- 425
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 358 ROMIYFPRKVKTLTKFDEHENGJFKHLLDAER---HQJOWK---HSPYF---YF 400
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 -----YSDSNIVALL---SHKKGATQCNIMHALEKTKKPPDSDIYMKKFEKJHFSQVSA 415
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 409 QIXSLHDY-IREIDRLSGDKNTAIYTFGQ---FRPEFDLIFRPAIGYOKALEPL 465
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 DLMANNAHDFITISTOYLQAGTKNTV-----GQYESHAFFPPLVYWHGIDVDPKFN 445
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 466 LRSPATKVIKTEINREMIETERFQGHGYHYILIMDIFKDLINVIAMOMTJAGT 525
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 466 IVSPGADWAIYFPESEK---DVTGUTSLHLHLEQLLFPBQNEHIOVLDNDISKPIFSM 525
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 526 DTIHHPDHYG 536
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 523 ARLDKRVKNTTG 533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
Search completed: October 12, 2001, 16:03:55
Job time: 78 sec
```

Fri Oct 12 16:09:18 2001

us-09-729-454-2.rai

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2001, 16:04:29 ; Search time 29.42 Seconds
(without alignments)
1416,298 Million cell updates/sec

Title: US-09-729-454-2
Perfect score: 2908
Sequence: 1 MSSEMLQKTLILISFSVY.....IHPDHYIGNIMFLNYIC 547

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2320	79.8	540	2	B45665	adult-specific 61.
2	119.5	4.1	860	2	T28227	ORF MSV067 probabl
3	117	4.0	792	2	T00082	hypothetical prote
4	112	3.9	938	2	A39160	transcription acti
5	104.5	3.6	1253	2	T45787	disease resistance
6	103.5	3.6	1237	2	S64385	probable membrane
7	103	3.5	801	2	T06783	ent-kaurene syntha
8	100	3.4	527	2	A81712	4-alpha-glucanota
9	100	3.4	698	1	S00742	Cytochrome-c oxida
10	99.5	3.4	400	2	S44810	F44B9.7 protein -
11	99.5	3.4	614	2	B43935	spat protein - Bac
12	98.5	3.4	368	2	S46435	sodium precursor -
13	98	3.4	699	2	S29499	probable membrane
14	98	3.4	749	2	S61643	DNA topoisomerase
15	98	3.4	761	1	B64506	trehalose-phosphat
16	98	3.4	817	2	T43659	probable membrane
17	98	3.4	892	2	S57055	thymidylate synthase
18	98	3.4	1025	2	T59311	exodeoxyribonuclease
19	98	3.4	1121	2	G64103	hypothetical prote
20	97.5	3.4	1049	2	T37953	hypothetical prote
21	97	3.3	675	2	S53831	NADH dehydrogenase
22	97	3.3	1324	2	T01508	mismatch repair en
23	96.5	3.3	671	2	D72346	chemotaxis sensor
24	96	3.3	1063	2	T38420	probable DNA helic
25	96	3.3	1103	2	A85189	disease resistance
26	95.5	3.3	823	2	T34472	hypothetical prote
27	95.5	3.3	1352	2	G84473	hypothetical prote
28	95	3.3	838	2	A48440	ring-infected eryt
29	95	3.3	1148	2	T18770	probable calcium c

30	95	3.3	1274	2	I40813	neurotoxin type F
31	95	3.3	1714	2	E71609	Ser/Thr protein ki
32	94.5	3.2	4981	2	T18489	hypothetical prote
33	94	3.2	476	2	A56510	dolichyl-diphospho
34	93.5	3.2	227	2	D81259	probable tonb tran
35	93	3.2	819	2	T07319	hypothetical prote
36	93	3.2	892	2	B36849	A10L protein - var
37	93	3.2	1852	2	T28552	hypothetical prote
38	93	3.2	1857	1	S01787	fatty-acid synthas
39	93	3.2	2467	1	D71437	probable resistanc
40	92.5	3.2	424	1	B64216	hemolysin tlyc hom
41	92.5	3.2	502	2	H83716	two-component sens
42	92.5	3.2	686	2	I49298	dystonia isoform 2
43	92.5	3.2	1049	2	S67613	vacuolar carboxype
44	92.5	3.2	2875	1	RYUTW	genome Polypeptide
45	92	3.2	891	1	WMY23W	major core protein

ALIGNMENTS

RESULT 1
B45665
Adult-specific 61.9K brush border protein precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 03-May-1994 *sequence_revision 03-May-1994 *text_change 04-Mar-2000
C:Accession: B45665
R:Boill, W.; Schmid-Chanda, T.; Semenza, G.; Mantel, N.
J. Biol. Chem. 268, 12901-12911, 1993
A:Title: Messenger RNAs expressed in intestine of adult but not baby rabbits. Isolati
A:Reference number: A45665; MUID:93286138
A:Accession: B45665
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1540 <BOL>
A:Cross-References: GB:212840; NID:91761; PIDN:CAA78302.1; PID:91762
A:Superfamily: rabbit adult-specific 61.9K brush border protein
C:Keywords: intestine

Query Match	79.8%	Score 2320	DB 2	Length 540:
Best Local Similarity	78.8%	Pred. No. 1.9e-176:		
Matches 427	Conservative 59	Mismatches 54	Indels 2	Gaps 2:
QY	6	MLQKTLILISFSVYTWMIPIISQNFKLMSALNISVHYMNNKSAKSLFKTSLIFLKP	65	
DB	1	MLRKYKLICLLAIC-VLCIISONSSTKIMGALKIPNSHYNTSMISSIFKMSVPSKS	59	
QY	66	LTELEIRKEITEIKIDQIPRPPTHTVNTTTSATSTATTINPDYVCGDQDILLEV	125	
DB	60	LTELRLRVEIKLEIKDRILIPRPPTHTVNTTTSATSTATTINPDYVCGDQDILLEV	119	
QY	126	DHLGOKRQVGFPLARMSPALTACASQKVMDFNNGTYLVSTFLMEQVSLSLILHP	185	
DB	120	DYLGHOKEVGFPLARMSPALTACASQKVMDFNNGTYLVSTFLMEQVSLSLILHP	179	
QY	186	SEGASALMRARNOGYDKIIFKGFVNGTSHVTECGTLTNSMLECYLDDRDQAEFYCM	245	
DB	180	SEGASALMRARNOGYDRILIFGQVNGTSHVTECGTLTNSMLECYLDDRDQAEFYCM	239	
QY	246	KPOMCEALTYWTRNREVSYLTDKENSLEFRSVCVEMMKRKIIDVTGNMKREIEE	305	
DB	240	KPOMCEALTYWTRNREVSYLTDKENSLEFRSVCVEMMKRKIIDVTGNMKREIEE	298	
QY	306	TCQVGMKPPVPGGYTLQGWITTCFNOVOLDTIKINGCLGKLIYLGDSLTRQWIIYFP	365	
DB	299	KCOIGMKIIPVPGGYTLQGWITTCFNOVOLDTIKINGCLGKLIYLGDSLTRQWIIYFP	358	
QY	366	KVYVTLKFPDLEHETGIFPKHLLDARRHOIQMKHSPVTFOLSLDHDYIPRPIED	425	
DB	359	KVYVTLKFPDLEHETGIFPKHLLDARRHOIQMKHSPVTFOLSLDHDYIPRPIED	418	

RESULT 2
T28k27
 ORF MSY067 probable mRNA capping enzyme large subunit homolog (vaccinia DIR) - Melanoplus sanguinipes entomopoxvirus
 C:Species: Melanoplus sanguinipes entomopoxvirus
 C:date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T28k27
 R:Alfonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
 J. Virol. 73, 533-552, 1999
 A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
 A:Reference number: Z20484; MUID:99102612
 A:Accession: T28k27
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-860 <APO>
 A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:ACG97615.1; PID:g4049855
 C:Genetics:
 A:Note: MSY067
 C:Superfamily: poxvirus mRNA capping enzyme large chain

QY 438 FGGHPPPIDIFRAHGVCAALRELFLESPATKVIITKENTENEMHILETGGDFHGY 477
 Db 568 NNQOORP-PLNIF-----TNLIKTEAISIASKLCAEIPNPIVATSIDIGGCHINNY 619
 QY 498 HVLIM-----KDI-----FKLWVGID----- 515
 Db 620 YIGISGMIGTDPDLSAEOERINSLQTKRRRNSTVFKFSLISLTDHLYLNKVKSTY 679
 QY 516 -----AMDNTI--AYGDTTHPRDHY-----GNCI--NPELN 544
 Db 680 MFOOKIKYFGLIEMQAHYHYSVNDTKNLINLNKSLSTDSGTKYITICLNQKKNKLLIN 739

[illegible]

A:Molecule type: DNA

A:Residues: 1-938 <DEH>

A:Cross-references: GB:M60105; NID:q143149; PIDN:AAA22572.1; PID:q143150

R.Kunst, F.; Ogasawara, N.; Koszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertier, C.; Bron, S.; Brouillet, S.; Bruchet, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Chd, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.; Nature 340, 249-256, 1997

A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, A.; Gallet

lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kutita, K.; Lapius, S.; Masuda, S.; Maueed

Y, M.; Ogawa, K.; Ogilwara, A.; Oudea, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Pottebell

Rieker, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schriener, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunori, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danclio, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MID:98044033

A:Accession: A69651

A>Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-938 <KUN>

A:Cross-references: GB:Z99117; GB:AL009126; NID:q2634966; PIDN:CAB14650.1; PID:q2635154

A:Experimental source: strain 168

C:Genetics:

A:Gene: levR

C:Superfamily: RNA polymerase sigma factor interaction domain homology

C:Keywords: DNA binding; nucleotide binding; P-loop; transcription regulation

E:117-344/Domain: RNA polymerase sigma factor interaction domain homology <SFI>

E:148-155/Region: nucleotide-binding motif A (P-loop)

Query Match 3.9%; Score 112; DB 2; Length 938;

Best Local Similarity 21.0%; Pred. No. 1.3; Indels 176; Gaps 38;

Matches 130; Conservative 100; Mismatches 213; Indels 176; Gaps 38;

QY 3 SNTMCKTLLILISFSVYVWTFIISQNTKLSALNLSI-SYVWYNSASLEPKTSLI 61
 DB 53 NNLVRSKVIYIKITFPRVPIVIAEKLEPKMKWDEMEVCKDLQAFSGSKONHOISTN 112
 QY 62 PLKPLETELRKIKETIEKLDQOI--PPRPFHVNTTTSATHTATILNPDYCKGQOLD 119
 DB 113 PLELMIGAKGSLKAKISQAAVFP--NGHL--MLLGTFS--GSL-- 156
 QY 120 LLEVMHLLGQKQYGDPLKAMSSPALTAGASQKVDENCTYLV-----SFT 169
 DB 157 ---FANRIYGFATY-SDILKA--GAPFITENCA---DYNNPQLLSQLGHHKGSFT 205
 QY 170 LFWGGVSL-----SILL---IH--PSGASALMRARNG--VDKIIFGKFN--GTSV 216
 DB 206 GADEKAGLVQDANGILFMDETHLRPFEGOBMLEFYIDSGSYNRL--GESEHKRTSNV 262
 QY 217 FTECGITLNSNAELCEYLIDRDQEAFCMKPO--HMP-----CEALTYMTNRE 264
 DB 263 LFIQATTEPSSALL-----KTELRIRPTHTIHISLEERSLNERVDTLTFILGKEAE 314
 QY 265 -----VSTLTUKENSLFHRKVG--VEMMKDR-----KHIDVTCNNKREKIEETQOVGMP 313
 DB 315 KIKKMLSVHIDVYVNLHSAKPFQNGQLKSNVQLCAHGLHLDNEVETLYR----- 369
 QY 314 PVPGYTLGGKWTTFGNO-----VOLDTIKINGCLNGKLTLLD 354
 DB 370 DLPD--EIKQEWSSSKNMORSKAISEVYNTITLISPIVEDETTKIDELSPNLYHLEE 427
 QY 375 STLQWIVYFPKVVTKLFPDLHETGIFKK---HLLDAERTQIQMKKSYFPYFOL 410
 DB 428 K-----VKTLM-----KEGLSKDINOITLIDVHLVRSF--HHQAQKDL 468
 QY 411 YSLIDHDYIPREDIRLSGDNTAIVTTFQO--HPRFPIDIFIRRAIOVKAIEFLR 467
 DB 469 LTFVHIMVI-QMKQKELIAHEHLDCTFDREKFIYFLSMHIDATLRKG-----KQIDVINTQ 523

QY 468 SPATKVIKTEKREHIER-----FGD-FRGY-----IHYLIM-----KQIFPD 508

DB 524 -----ETDEIDHTVHKERYVAMIFKDKIQEYFKAIPDEIVLYITMLHSTIKSLKEN 575

QY 509 LNVGIIDAWMTIYAGTDT 527

DB 576 KRVGIIVA-----AHGNST 589

RESULT 5

745787 disease resistance-like protein - Arabidopsis thaliana

N:Alternate names: protein F26013.200

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T45787

R:Delserny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laude, M.; Mewes, H.W.; Lemcke,

submitted to the Protein Sequence Database, December 1999

A:Reference number: 223013

A:Accession: T45787

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1253

A:Cross-references: EMBL:AL133452

A:Experimental source: cultivar Columbia; BAC clone F26013

C:Genetics:

A:Map position: 3

A:Insertions: 158/2; 490/3; 589/3; 1047/3

A>Note: F26013.200

Query Match 3.6%; Score 104.5; DB 2; Length 1253;

Best Local Similarity 17.5%; Pred. No. 7.6; Indels 247; Gaps 27;

Matches 108; Conservative 78; Mismatches 183; Indels 247; Gaps 27;

QY 42 ISVHYNSAKSLPKTSLPLKP-LTELELRKIKETIEKLDQIIPRPFTVNTTTSATN 100
 DB 230 LEAHFGKILREELGKSSI--TRPILRLVLRHRYLVLDVCKPL----- 274
 QY 101 STATILNPDYCKGQD-----QLDILEVD-----HLGQKQGGDF 138
 DB 275 DASPLGQDMFCGSLIITTSKDKVNSICRVQIIEYVGLNEBALQLSRKARKEI 334
 QY 139 LRAKSSPALTAGASQKVDENCTYLVSTLFWGVSL-----LLIHPSG--ASA 191
 DB 335 IHPSLCK-----LSKRVIDYANGNL--ALIFGGMSSKKNKPIELAFKVKYLAHE 385
 QY 192 LMRARNGVDKI-----IFGKRVNGTSHVTEG--LTLNSNAELC--- 231
 DB 386 IHAYVSTYDLSLSNEKNITPLDIACLFQGENVCYITLLEGGCFPRVEIVLVKCLVS 445
 QY 232 -----EYDDR-----DOEAFYC----- 244
 DB 446 MAEGRYVMHNLQISGRKIINGKRRSRMLPLIKKFELEDROVLGSEDIAIFLDPSAL 505
 QY 245 ---MKP-----OHMPCEALTY- 257
 DB 506 SEVDNMAPENNYMLRYLKTCSSPNCKYALHLKQVKSILPELRILMHEIFLLSLPD 565
 QY 258 MTRNREVSYLTDKENSILFHRKVGVEYMKDKRHIDVTCNNKREKIEETQOVGKRPV 317
 DB 566 FNTNRNLVILNMC---YSKLRQRLWEGKELGKIMKIMCHSOQLVGIOE-LQIALNMEV-- 619
 QY 318 GYTLGCKWTFPCNOVO-----LDTIKINGCLGKGLYLLGDSITLQWITYPVPV 367
 DB 620 -IDLG-----CARLQRFATGCHQHLRVNLNLSGCIK-----IKSPFEY 657
 QY 368 VKTLKFFEDLHETGIFKKHLLDAERTQIQMKKSYFPYFQ-----LYSLIDHDYIPR 421
 DB 658 PRNIEELVLYKQGI-----BSIPTVYFSPQNSFIYDKDKHFLNR 698
 QY 422 EIDRLSGDNTAIVTGFQHFRFPID--IFIRRAIOVKAIEFLRLSPATK----- 472

Db 699 EV---SSDSOSLSIMVYLDNLKVLDSLQCELEDEDIGIFPKRLKLYLGTAIKEPLSLMH 755
 QY 473 ----VIKTENIREMH 484
 Db 756 LSELVYLDLENCKRLH 771

RESULT 6

Probable membrane protein YGR090W - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein G4639
 C:Species: *Saccharomyces cerevisiae*
 C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 20-Jun-2000
 C:Accession: S64385; S64395
 R:Medler, H.; Scharfe, M.; Medler, E.; Mamont, R.
 Submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64386
 A:Accession: S64385
 A:Molecule type: DNA
 A:Residues: 1-1237 <MED>
 A:Cross-references: EMBL:Z72875; NID:91323132; PID:91323133; MIPS:YGR090W
 A:Experimental source: strain S286C
 R:Hernandez, K.; Weber, N.; Wipfl, P.; Schmidheini, T.
 Submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64392
 A:Accession: S64395
 A:Molecule type: DNA
 A:Residues: 1-1237 <HER>
 A:Cross-references: EMBL:Z72875; NID:91323132; PID:91323133; MIPS:YGR090W
 A:Experimental source: strain S286C
 C:Genetics:
 A:Map position: 7R
 C:Superfamily: *Saccharomyces cerevisiae* probable membrane protein YGR090W
 C:Keywords: transmembrane protein
 F:385-401/Domain: transmembrane #status predicted <TM>

Query Match 3.6%, Score 103.5; DB 2; Length 1237;
 Best Local Similarity 19.3%, Pred. No. 9;
 Matches 102; Conservative 87; Mismatches 175; Indels 165; Gaps 28;

QY 30 NFKTMSALNLSISVHYMNSAKSLFPKSLIPKLTETLRKEITELKDOQIPPRP 89
 Db 284 NFKTRRSINLLGFPKYFEKPKLLPNRNCI-----RIAQ--EKEKQSLPAPPL 331
 QY 90 THVNTTSATHTATILNPRDTYCRGDQDILLEVHDLGQRKQYGGDFLRARM----- 143
 Db 332 YNFSSLSSTH-----ENYLT-----YLYTKKQTESFEVATVIGRLML 370
 QY 144 -----SSPALTAGASGVMDFNNGTYLSFTLFW-----GVVSLSLLIHSEGASLM 193
 Db 371 QQRGFSSMSHSGSLG-----GFGTF--EFTTLMALLNGGGINSKNLILH----- 414
 QY 194 RARNQGYDKI-IFKG--KFWNGTSHVTECGTLNSNAE-----LCEYLDDRQDEAYC 244
 Db 415 -----GESSYQLKGVYKYL-ATMDLCHDGLQFHSNENSSSPASKYIDGFGPTPLF 468
 QY 245 MKPQHPCEALTYWTRNREV--SYLTDKENSLEFRSKVGVEMKKOR-KHIVTGNKRE 301
 Db 469 DKSTKV--NLTKNITVSQYLKEXAGETIRMLNN-----VQDDFSNITLTIISFED 519
 QY 302 KIEETQGVKMPVPGGTYLQCKWITTFQNOVQDITIKNGCLKGLTYLLDSTLRQMI 361
 Db 520 NLK-----YDL-----CYDQDLPGRYNN-LETSIAATFSGMERVKFI 556
 QY 362 YFPRVVKTLKFPDLHETGIFKKHLLDAERHTQIQW--KKHSYFPVTFOLTS-----LI 414
 Db 557 -----TLENFLAKITINARVALGDRIKYIOLEVMGQASDPRTKKKVSNGNHMF 608
 QY 415 DHDYI-----PREIDRL--SGDKNATVITFGHFRFPIDITFIRAKVQAKIER-- 463
 Db 609 NDFEVRYKLIYNPSECCLKLVTKGPAHSETKSTEAIVKNF-----WGIKSSLRRFK 659

QY 464 -----LELRSPATKVIKTENI-----KEMHETERFGFQCTI 497
 Db 660 DQSTIRHCCVWSTSSSEPTISSTIVNALQKHVSKKAOISMETIKRHNEL 708

RESULT 7

ent-kaurene synthase A (EC 2.5.1.-) - garden pea
 C:Species: *Pisum sativum* (garden pea)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: T06783
 R:Alt-All, T.; Swain, S.M.; Reid, J.B.; Sun, T.P.; Kamiya, Y.
 Plant J. 11, 443-454, 1997
 A>Title: The *us* locus of pea encodes the gibberellin biosynthesis enzyme ent-kaurene synthase
 A:Reference number: Z15810; MUID:97260958
 A:Accession: T06783
 A>Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-801 <AIT>
 A:Cross-references: EMBL:U63652; NID:92160543; PID:UAB58622.1; PID:92160544
 A:Experimental source: cultivar Torsdag
 C:Genetics:
 A:Gene: *us*
 C:Function:
 A:Description: converts the geranylgeranyl diphosphate into copalyl diphosphate
 A:Pathway: the first step of the gibberellin biosynthesis
 C:Keywords: transferase

Query Match 3.5%, Score 103; DB 2; Length 801;
 Best Local Similarity 19.8%, Pred. No. 5.3;
 Matches 124; Conservative 91; Mismatches 197; Indels 214; Gaps 14;

QY 2 SSNTMLQKTLILISFVVTW-----MFIISQNPFKLMSALNLSISVHYMNSA 1
 Db 154 SAHRIINTLACVIALR--SNMNMSECKDGMF--FRENLSKLENEHEHMPGF----- 205
 QY 52 KSLPFRKSLP-----LKP-----TELRRIKETEKKDOIP--PFRPHVNTTSATHT 102
 Db 206 EVAPR--SLLEGARQIKPLMCPNDSPI-LKNIFERDKLRIKELIMHNVPT----- 265
 QY 103 ATILNPRDTYCRGDQDIL-LEVHDLGQRKQYGGDFLRAMSSFPALTATASAKVMPNN 161
 Db 256 -TLHSLEGSGHDMQKQLKQSD-----GSFL-----FSPSTANALMQIKLQNG 401
 QY 162 GTLYVSFTLFWEGVSLSLLIHSEGASLM--RARNQGYDKIIFKGF--VNGTSH 215
 Db 302 LKTLNNVYKKFNGVPR--NVYFVDLFELIHWVDRLRLGISR-FRHEIKDCMNVYK 355
 QY 216 VTECGI--TLNSNAELCEYLDROEA-----FYCKRP 247
 Db 357 IMSEKGIQVARNSNVQ-----DIDTAMAFRLRLRHGHQVSAHVFRHFRNGEPPCFAG 410
 QY 248 QHMPG--EALTYWTRNREVSYLTDKENSLEFRSKVGVEMKKDRH----- 291
 Db 411 Q--CIQAVTGMNLFRAVOVLPFGKEILHAKHSAKALEKKEANELLQWIMIKSLP 467
 QY 292 -----IDVTGNKREKTEETG--QVKKMPVPGGTYLQCKWITTFQNOVQDITIKNG 344
 Db 468 EEVGYALDPMWYANLRIETRFYIDYGASQVWIGTKLIMAVANNVNNVLLARLQYNN 527
 QY 343 CLKGLTYLLDSTLRQMIYFPRVVKTLKFPD-LHETGIFKKHLLDAERHTQIQWKKH 401
 Db 528 CQAGHL-----EW-----NVQIOTYLESRLGEFGLSKROLL----- 560
 QY 402 SYPRVYFQKLSLDHDIYFRELDSGDKNATVITFGHFRFPIDITFIRAKVQAKAI 461
 Db 561 AVPLATGSIFF-----PERSHERLMAATTALETI----- 591
 QY 462 ERLFSPATKVIKTENIREMHETERFGFQHYIYLLMKDIFPD-----LN 510
 Db 592 -----KCYVNEDELKRPD--AKKNCHIDVDYQSIARBMKMKTEHIVESLFIAT 639

OY 511 VGIIDAMDTIAYGTDTIHPPDHVIG 536
 Db 640 IGET-SMDVRLSYG-----HEIG 656

RESULT 8

4-alpha-glucanotransferase TCO362 [Imported] - Chlamydia muridarum (strain N199)
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
 C:Accession: A81712
 R:Read, T.D.; Brundham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255
 A:Accession: A81712
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-527 <TET>
 A:Cross-references: GB:AE002303; GB:AE002160; NID:g7190392; PIDN:AAF39223.1; PID:g719040
 A:Experimental source: strain N199 (Mopn)
 C:Genetics:
 A:Gene: TCO362
 C:Superfamily: 4-alpha-glucanotransferase

Query Match 3.4%; Score 100; DB 2; Length 527;
 Best Local Similarity 20.0%; Pred. No. 5.1;
 Matches 92; Conservative 68; Mismatches 158; Indels 142; Gaps 25;

OY 65 PLETETLR--KEIIKILDOOI--PPR-----PETHVTTTSATHTATILNPRDTCR- 114
 Db 2 PLTSRLRIQNSPIKRWYQVDTSPKHGICVPLFSHTQNSGIGELFDILPMIDCTL 61
 OY 115 -GQDDILEVRDHLGQRKQYGDPLRAMSSPALTAGASGKVMNNGTYVSFTLME 173
 Db 62 GGFQILQILPINDTSCSSPPNS-----ISISL-----NPLHLISALPYK 103
 OY 174 GQVSLSLILHPSEASALWRARNOGYDKII-FKGFVNQTSHTFEG-GTLINSNAELC 231
 Db 104 EEVSSSRKLIQEMQRLSDSQV---NKEKVIPIKRAFF---KEKPYCKSKNLNHRDFC 157
 OY 232 EYLDKDDQ-----EAFYCKMPQMPCEALITYTRNNEVSYLTDKNSL-----FHRS 279
 Db 158 DFC-EKEKYLHPYALFCSIREHLNLPINMSSTTYDLSYISQHEHIFANDIEFYSYLD 216
 OY 280 KVCVENMKD-RKHIDVTNCKNREKI-----EETCOV-----GKRPVPGGYTLQ 322
 Db 217 YLFFEQMKQVKKHADHKGLIKGDIPILISKDSCDVFYKRYSSSESYSQSP-PDFYNAE 275
 OY 323 GK-WITTCNOVOYDITKINGCLKGLIYLLGDSITLQOMITYYFPKVKYTKL----FPDL- 376
 Db 276 GQMNMLPIYNN-----FTLQADAYHMMK--EELRYAENFYSLY 311
 OY 377 ---HETGIFKKHLLDAERHTOIOMKKHSYFVTFOLYSIDH-----DIYPREID 424
 Db 312 RLDPVVGLEF-----FWWDELGGRFPDPPDPRDLDDGTD 347
 OY 425 RLSCGKNTAIVTFGOHFRFPIDIFIRRAIGYOKAIEL 464
 Db 348 ILSHLKASSKMLPIGEDLTTPVD-----VQADLESL 379

RESULT 9

S00742
 C:Species: cytochrome-c oxidase (EC 1.9.3.1) chain I - Tetrahymena pyriformis mitochondrion
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999
 C:Accession: S00742
 R:Zlate, Z.; Suyama, Y.
 Curr. Genet. 12, 357-368, 1987
 A:Title: The cytochrome oxidase subunit I gene of Tetrahymena: a 57 amino acid NH2-termi

A:Reference number: S00742; MUID:88184706
 A:Accession: S00742

A:Molecule type: DNA
 A:Residues: 1-698 <ZIA>
 A:Cross-references: EXBL:X06133

C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGC6

A:Start codon: ATA
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane
 F:59-617/Domains: cytochrome-c oxidase chain I homology <COI>
 F:111,538/Binding site: heme a iron (His) (axial ligands) #status predicted
 F:401,450/Binding site: copper (His) #status predicted
 F:401-405/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F:405/Binding site: oxygen (Tyr) #status predicted
 F:528/Binding site: magnesium (His) (shared with chain II) #status predicted
 F:536/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 3.4%; Score 100; DB 1; Length 698;
 Best Local Similarity 19.2%; Pred. No. 7.5;
 Matches 105; Conservative 81; Mismatches 194; Indels 166; Gaps 30;

OY 23 MIFITSONFTKLM-----SALNLSIVH---YWNNSAKSLPPTSLIPKPLTELRIRK 74
 Db 1 MLVTFN-SFDNNMWFDEIQTSPKRYSVNNYVYLNKIKLFL--TYLNDR-----K 48
 OY 75 EIEIKDQOIPPRPF-----THVTTTSATHTATILNPRDTCRQDDILLEY 124
 Db 49 HILKKVYVYINNKRIAINLYFSWTLGSAALATMIRMLAPHSFPFKDLSRLYQV 108
 OY 125 KDHLGQRKQYGDPLRAMSSPALTAGASGKVMFNNGTYLVSTL----FNE--GGV 176
 Db 109 TAH-----GLIMFFVVPVPLTFEGFANFLIPYHVSQDAVYPRLSIGFWIOPGYI 160
 OY 177 SLSL-LIHPSEASALWRARNOGYDKII-FKGFVNQTSHTFEGGLTINSNAELCEYLD 235
 Db 161 LKAKIFLPO-----FWRV---YKTSFSPLEKMK-----NOKRYKN 199
 OY 236 D-----KQCAFY-----CKPQH-----MPEALITYTRNNEVSYLTQXE 272
 Db 200 DYLFYDLKKEITDDHSFFWKARKYIKPOYSVFSPLKIMMKMTMINPESF----- 254
 OY 273 NSLFHRSKYGVENMKDRKHIDVTNCKNREKIEETCOVGMKRPVPGGYTLQGMWITTCNO 332
 Db 255 --WYASRY--VQSRKKYVYKCSARTL-----TTAGT-----FITPSS- 292
 OY 333 VQDITKINGCLKGLIYLLGDSITLQOMITYYFPKVKYTKL--FPDLHETGIFKKHLLDAER 392
 Db 293 ---NRTYTG-----VGSODILILSVFAGISTTISFTNLL--ITRRLTAMGKM 336
 OY 393 HTQIOMKKHSYFVTFOLYSIDHDIYPREIDRLSGKNTAIV-TFQOHFRFPIDIFI 451
 Db 337 HRRV-----LMPFVTISIFLTL-----RMLATITPVLAAGVYMAVAFRRHMQ---TTFF 381
 OY 452 RRAIGYOKAIEL--LFLRSPAKYIKTE---NINEMHIEIFERPGFH-----G 495
 Db 382 EYAVGDDPLISQHLFFEGHPEVYVLTIPTFGFINMIVPNNNTRVASKHMMIAYVMA 441
 OY 496 YHYLL 501
 Db 442 YMGYLV 447

RESULT 10

S44810
 R4489.7 protein - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
 C:Accession: S44810
 R:Antonaci, Pulton, L.

Submitted to the EMBL Data Library, September 1993

A:Description: Sequence of the C. elegans cosmid F4489.

A:Reference number: S44807

A:Accession: S44810

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-400 <NNT>

A:Cross-references: EMBL:L23648; NID:g388585; PID:g388589

C:Genetics:

A:Introns: 49/2; 124/3; 188/3; 233/2; 274/2; 321/1; 358/2

Query Match 3.48; Score 99.5; DB 2; Length 400;

Best Local Similarity 22.08; Pred. No. 3.7;

Matches 70; Conservative 38; Mismatches 129; Indels 81; Gaps 15;

```

QY 56 PKTSLI-----PLKPLTELEIKELIKEDQOIPRPETHVNTTSATSTATI 105
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 62 PTTSLAQHLSSAPLPAPPOQSHL-----SQLDQPTTSTATNPHAPITOTI 111
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 106 LNPRTYCRGDODILLEVRDHLQKQYGGDFLARMSALTASGVKWDENNGTYL 165
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 112 ARPDNARNADQDELAAKQ--QOQEX--MLQOHQOQONRAQQLAQEAFAN---- 162
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 166 VSTLFMEGVSLSLILHPSEGASALMARNOGYDKIIFKGFVNGTSHV----FTFCG 221
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 163 -----PPTVSIPQTGHPPQTSASQQAQDE-----TMAQGTPEQSHIESPKETN 210
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 222 LTLSNAELC---EYLDROQAEFYC-----MKPQIM-----PCE-ALTYM 258
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 211 LEKSTGELCLYGRVNDLVNKTSTYSLTKKVERKPLQGENPNDLADCKAALQRM 270
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 259 TTRNREVSYLTDKENSLEFRSKVG---VEMKD---RKHLDTYCNKKREKIEE--TCQV 309
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 271 S---DIRQIEKRREPTMKRMTGEDYIELMDSELKKPWDEBQKRAELEQFAVI 326
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 310 GMRPVPVGGYTLQKMIT 327
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 327 SKTPVEGHVWGKMS 344
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 11

B43935

Spot protein - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 10-Mar-1993 #sequence revision 18-Nov-1994 #text-change 02-Feb-2001

C:Accession: B43935; B42655; I40512; I39978

R:Klein, C.; Kaletta, C.; Schnell, N.; Entian, K.D.

Appl. Environ. Microbiol. 58, 132-142, 1992

A:Title: Analysis of genes involved in biosynthesis of the lantibiotic subtilin.

A:Reference number: A43935; MUID:92171481

A:Accession: B43935

A:Molecule type: DNA

A:Residues: 1-614 <KLE>

A:Cross-references: GB:M86869; NID:g143713; PIDN:AAA22838.1; PID:g143715

A:Experimental source: ATCC 6633

A>Note: sequence extracted from NCBI backbone (NCBIN:84011, NCBI:P:84013)

R:Chung, Y.J.; Steen, M.T.; Hansen, J.N.

J. Bacteriol. 174, 1417-1422, 1992

A:Title: The subtilin gene of Bacillus subtilis ATCC 6633 is encoded in an operon that

A:Reference number: A42655; MUID:92138640

A:Accession: B42655

A:Molecule type: DNA

A:Residues: 16-38 'A', 40-67 'D', 69-93 'S', 95-106 'Q', 109-614 <CHU>

A:Cross-references: GB:M83944; NID:g143557; PIDN:AAA22770.1; PID:g143559

A:Experimental source: ATCC 6633

A>Note: sequence extracted from NCBI backbone (NCBIN:79666, NCBI:P:79668)

R:Klein, C.; Entian, K.D.

Appl. Environ. Microbiol. 60, 2793-2801, 1994

A:Title: Genes involved in self-protection against the lantibiotic subtilin produced by

A:Reference number: I40511; MUID:94368094

A:Accession: I40512

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-614 <RES>

A:Cross-references: EMBL:U09819; NID:g2702240; PIDN:AAB01567.1; PID:g595.1

C:Genetics:

A:Gene: spaB

C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette

C:Keywords: antibiotic biosynthesis; ATP; nucleotide binding; F-loop

F:370-569/Domain: ATP-binding cassette homology <ABC>

F:387-394/Region: nucleotide-binding motif A (P-loop)

Query Match 3.48; Score 99.5; DB 2; Length 614;

Best Local Similarity 20.48; Pred. No. 6.9;

Matches 128; Conservative 88; Mismatches 212; Indels 19; Gaps 29;

```

QY 5 TMQKTLILISFVYTMFIIS-----QNFK 34
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 28 TLERSFLKILFSLITIGLIVSLYISOELINSLVITREVSIVITFLYLVNPFSE 87
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 34 LWS-----ALNLSIVHYWNSAKSLFPKTSLLPKPLTELEIKELIKEDQO 83
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 88 LISQISEFYNGKFOILIGYKLVN-----KYKKKSNLALKDFENPEIYDK--LERYTKE 139
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 84 IPRPETHVNTTSATSTATIL-----NPRDYGQDGLDILLEVRD----- 126
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 140 ISKRPYQIQAIIITTSFVTLSTIAFLMSKNR-----VSLLVIVISLFYFL 191
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 127 HLGQRQY-----GGDFLARMSSPALTAGASGVMP--FNGTYLVSTFLMEGVSL- 178
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 192 KIQDEEFTFMKRAKERRKSMYSIYLTIDFSFELKMLYMLKDYLLN--KYWDIKKSFIE 249
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 179 -----SLLIHPSE-----GASALMARNOGYDKIIFKGFV-----NCTS 214
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 250 QDTKILRKTLNLIELAVQVAGVYIIIALMSFAGKIMVGNVMSYRSVLYQHSQ 309
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 215 HVTEEGGLTNSN---ELCEYLDNDQDAFYCKPQMPCELYITTRNREVSYLTK 271
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 310 SIMTSYISLYNSNLYMNQLYEFLLEKESSQCHKPVEPIHNVF---QNVSEFYVNG 366
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 272 ENSLFKRSKVGEMKDKKHIDVTCNCKREKIEETCGVGMKPPVPGYTLQCK---WIT 327
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 367 EQLT-----KHIVNS-----LHNGENAVLG-----PNS--SKKTFIKLL 409
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 328 TPCNOVQDITIKNGCLKGLIYLLDSTLQWIIYPPKVK---TLK-----FEDLH 377
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 401 TGLYEHEGDLILINGI---NIKELMDSYMNQIALFQDFMKYEMTELENIGFQJDKIH 427
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 378 ETGIFKHHLLDAER-----HTQI-----QMKHSYFETPOLYS 412
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 458 QTN--KMEHVDIVRADFLKSHSSYQFQGLWFMFDEGRQLSGQWOKIALARVFEAS 515
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 413 LIDHDYIPREDIRLSGKNATAYITFGQHF--REPPIDITIRAGVQKALEPLASPA 470
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 516 LYLIDPSSALDPFAERE-----TFDFFSLSKDKIGIFISHRLVAALKRILVMDK 569
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 471 TKV-----IKTENIRENHIETEF 490
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 570 EIVGIGTHBELKTCPLKMKDSENY 596
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 12

S46435

Porin precursor - Fusobacterium nucleatum

C:Species: Fusobacterium nucleatum

C>Date: 14-Jul-1995 #sequence revision 21-Jul-1995 #text-change 17-Nov-2000

C:Accession: S46435; S51507

R:Boistead, A.I.; Tomassen, J.; Jensen, H.B.

Mol. Gen. Genet. 244, 104-110, 1994

A:Title: Sequence variability of the 40-kDa outer membrane proteins of Fusobacterium

A:Reference number: S46435; MUID:94316187

A:Accession: S46435

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-368 <RC1>
 A:Cross-references: EMBL:X72582
 R. Bolstad, A.I.
 submitted to the EMBL Data Library, August 1994
 A:Reference number: S51507
 A:Accession: S51507
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-156 / 'K', 158-368 <BO2>
 A:Cross-references: EMBL:X72582; NID:g530295; PIDN:CAA51172.1; PID:g530296

Query Match 3.4%; Score 98.5; DB 2; Length 368;
 Best Local Similarity 21.5%; Pred. No. 4;
 Matches 67; Conservative 35; Mismatches 132; Indels 77; Gaps 15;

```

QY 39 NLSISVHY-WNNSAKSLFPKTSILPLKPLETELKIKETLDOOIPRPPTHVNTTTS 97
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 57 NGSDVQYRMVGNVENRTPKKE-DPASPMLGDNVAGRL-----QTLTVNFTKQTLLEI 110
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 98 ATNSTATILNPRDTYCRGDQDLI-----LLEVRDLGOKROYGDFLARMS 145
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 111 RIRNYHTLMNPKDSQADDQYVHVFYKFKGLSSKIDVTSRL-EYKQNNQDAGKQAE 169
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 146 PALTAGASGVDFNNKGYLVFTLFWEGQVLSLLIHPEGASALMPARNQG-----Y 200
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 170 SVL-----PFDADYISNNFKADKFGFKL-----GYOKHMGHNSGVVQGP 212
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 201 DKILFKGFVNGTSHVFTEGTLTNSNA-----ELCEYLDDDRQDAFY----CKMK 247
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 213 NKGTQUNYFINFESEYTLPMWGFSAELNAVYNNKFKATYKNGKKGQFGELEATLY 272
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 248 QHMPCEALTYMTRRREVSYYTLTKENS-----LFHRSKGVGVEMKDKRHIDVNCNKR 302
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 273 QHTP-----LYKTNNVELSF--DEEGDYPTWTHQYK-----VSARDSNKEEYV 315
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 303 IEEFCOVGNKP 313
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 316 MLPTLQVSYKP 326

```

RESULT 13

S29499
 sodium channel protein alpha chain, epithelial, amiloride-sensitive - rat

C:Species: Rattus norvegicus (Norway rat)
 C:Date: 13-Jan-1995 #sequence-revision 13-Jan-1995 #text-change 09-Jun-2000
 C:Accession: S29499; S43503; S29715; S41158
 R:Linqueglia, E.; Voilley, N.; Waldmann, R.; Lazdunski, M.; Barbry, P.
 FEBS Lett. 318, 95-99, 1993

A:Title: Expression cloning of an epithelial amiloride-sensitive Na(+) channel. A new ch
 A:Reference number: S29499; MUID:93170495

A:Molecule type: mRNA
 A:Residues: 1-699 <LIN>

A:Cross-references: EMBL:X70521; NID:g433909; PIDN:CAA49916.1; PID:g433910
 R:Canessa, C.M.
 submitted to the EMBL Data Library, March 1994

A:Reference number: S43503

A:Molecule type: mRNA
 A:Residues: 2-598, DV, 601-699 <CAN>

A:Cross-references: EMBL:X70497; NID:q458845; PIDN:CAA49905.1; PID:g458846
 R:Canessa, C.M.; Horisberger, J.D.; Rossier, B.C.
 Nature 361, 467-470, 1993

A:Title: Epithelial sodium channel related to proteins involved in neurodegeneration.
 A:Reference number: S29715; MUID:93156815

A:Molecule type: DNA
 A:Residues: 2-194, P, 196-230, 'GAA', 234, 'LPATATT', 243-598, 'DV', 601-699 <CAN>

A:Cross-references: EMBL:X70497
 R:Canessa, C.M.; Schild, L.; Buell, G.; Thorens, B.; Gautschi, I.; Horisberger, J.D.; Rd
 Nature 367, 463-467, 1994
 A:Title: Amiloride-sensitive epithelial Na(+) channel is made of three homologous subun

A:Reference number: S41158; MUID:94150624
 A:Contents: annotation
 C:Superfamily: human amiloride-sensitive sodium channel protein; fibronectin type I
 C:Keywords: glycoprotein; sodium channel; transmembrane protein
 F:110-162/Domain: transmembrane #status predicted <TM1>
 F:422-457/Domain: fibronectin type I repeat homology <IPR>
 F:567-613/Domain: transmembrane #status predicted <TM2>
 F:425,539/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 3.4%; Score 98; DB 2; Length 699;
 Best Local Similarity 19.2%; Pred. No. 11;
 Matches 89; Conservative 53; Mismatches 157; Indels 164; Gaps 16;

```

QY 11 LLLISVYVMTIITISQNTKLSALSLISVYNNNSAKSLFPKTSILPLPLETE 70
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 115 VLMCTFGMMYQFALLFEELYSLPSLNIN-----NSDKLVPAVTCVLPNRYTE 168
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 71 LRKEIIEKLDO-----QIPRP 89
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 169 --IKEELFELDRITQTLFDLYKNSSYTRQAGARRSSRDLLGAFPHPLQRLTRPPPY 226
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 90 THVNTTTSATHTATILNPR-----DTYRGDQDLILEVRDH 127
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 227 S--GRARSGSSVBDNNDPOYDRKDKIGFOLCNQNSDCFYOTYSSG--VDAYRE---- 278
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 128 LGQKROYGCDPLFRAMSSPALTAGASGVDFNNKGYLVFTLFWEGQVSLILHPE 187
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 279 -WYFHYINILRLSDTSPALFEALG-----NFTICRFNPAPCNQANS-KPHHPMY 330
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 188 G-----ASALMPARNQGYDKITFKGFYNGTSHVFTEGTLTNSNAELCEYLD 238
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 331 GNCYTFENDKNSNLMWSSMPG-----VNN-----GLSTLTLEQNDFT---- 368
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 239 QEAFCYMKPQHPCEALTYMTRRREVSYYTLTKENS--LFHRSKGVGVEMKDKRHIDVNCN 298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 369 -----PLSTVTGARVYVHQQDPAFMDQDGNLRPEVETISMKREA--LDLSGN 418
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 299 KREKIEFCOVGNKPPVGGTYLQGNHITTFPNQVOLDITKINGCLKGLTYLGDSTLR 358
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 419 YGDCETENGSDVPVKNLYPSKITYQ-----VCHISCPQENMIKKCGCA--- 460
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 359 QMIYV-FPRVYKTLNF-----FDLHETGIFKK 384
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 461 -YIYPRKPGVEFCQYRKQSGWGYCYKLGAFSLSDLGCFSK 502

```

RESULT 14

S61643
 Probable membrane protein YOR081c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein 02964; hypothetical protein YOR2964c
 C:Species: Saccharomyces cerevisiae
 C:Date: 09-Mar-1996 #sequence-revision 12-Apr-1996 #text-change 21-Nov-1997
 C:Accession: S61643; S66966; S66964

R:Benes, V.; Andrade, M.A.; Reichmann, S.; Teodoru, C.; Banerji, A.; Sander, C.; Vale
 submitted to the EMBL Data Library, December 1995

A:Description: Nucleotide sequence and analysis of a 130 kb fragment of yeast chromo
 A:Reference number: S61643

A:Molecule type: DNA
 A:Residues: 1-749 <BEN>

A:Cross-references: EMBL:X94335; NID:g1262139; PID:e217721; PID:g1164929
 R:Voss, H.; Benes, V.; Reichmann, S.; Teodoru, C.; Schwager, C.; Paces, V.; Ansoyge,
 submitted to the Protein Sequence Database, July 1996

A:Reference number: S66965

A:Molecule type: DNA
 A:Residues: 1-749 <VOS>

A:Cross-references: EMBL:Z74989; NID:g1420242; PID:e251992; PID:g1420243; MIPS:YOR080
 A:Experimental source: strain S288C
 R:Bohn, C.; Bolotin-Fukuhara, M.; Dalgman-Fornier, B.; Dang, D.V.; Valens, M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S66929

Fri Oct 12 16:09:18 2001

us-09-729-454-2.rpr

Page 9

Db	120	DTLGNQKEVGDPLDARHMFSPALKAGASCKVTDENNQTYLVSFTLFWMGQVSLTVLLHP	179
Qy	186	SGSASALMRARNGDKEIIFKGFVNGVSHVFTGCGTILNSNAELCEFLDDROQEAFTCM	245
Db	180	SGSASALMRARNGDRIIFEGGVNGVSHVFTGCGTILNSNTECKLNGRQDFVFCM	239
Qy	246	KFOHNPCEALTYVTRNREVSYLDEKNSLFHRSKSVVEMMDRKHIDVYCNCKREKIEE	305
Db	240	KFOHNPCEALTYVTRNREVSYLDEKNSLFHRSKSVVEMMDRKHIDVYCNCKREKIEE	298
Qy	306	TOGVGMPPVPGSYLGGKMTTFPCNOYDPTINCKGKLYLLGDSNCKOMVYFP	365
Db	299	KOIGIMKIPVPGSYLGGKMTTFPCNOYDPTINCKGKLYLLGDSNCKOMVYFP	358
Qy	366	KVYKTLKFDLHETGIFKKHLLDAERHTQIQMKSHSPVTFQVLSLDHYIPEIDR	425
Db	359	KVYKTLKFDLHETGIFKKHLLDAERHTQIQMKSHSPVTFQVLSLDHYIPEIDR	418
Qy	426	LSGDKNTAIVIFGQHPFPFIDIFIRRAIGVKAISRLFLSPATKVIYKTEINREMI	485
Db	419	LTGKRDVYIYTFGQHPFPFIDIFIRRAISVQAIDRLSPATKVIYKTEINREMI	478
Qy	486	ETERRGFGHYIHYILMDIPKDLNANGIIDAMDVTNGDPDHPDVGNOINMFNT	545
Db	479	ETERRGFGHYIHYILMDIPKDLNANGIIDAMDVTNGDPDHPDVGNOINMFNT	538
Qy	546	IC 547	
Db	539	IC 540	
RESULT 2			
ID	LEVR_BACSU	STANDARD:	PRT: 938 AA.
IC	P23914:		
DT	01-MAR-1992 (Rel. 21, Created)		
DT	01-MAR-1992 (Rel. 21, Last sequence update)		
DT	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	TRANSCRIPTIONAL REGULATORY PROTEIN LEVR.		
GN	LEVR.		
OS	Bacillus subtilis.		
OS	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OX	Bacillus/Staphylococcus group; Bacillus.		
OX	NCBI_TaxID=1423;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=91172782; PubMed=1900939;		
RT	Debarbouille M., Martin-Verstetete I., Klier A., Rapoport G.;		
RT	"The transcriptional regulator Levr of Bacillus subtilis has domains		
RT	homologous to both sigma 54- and phosphotransferase system-dependent		
RT	regulators".		
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2212-2216(1991).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=168;		
RX	MEDLINE=97286548; PubMed=9141695;		
RA	Pardo V., San Roman M., Galindo I., Fumelle B., Bojotin A.,		
RA	Sotolin A., Meliádo R.P.;		
RA	"A 23911 bp region of the Bacillus subtilis genome comprising genes		
RT	located upstream and downstream of the lev operon.";		
RL	Microbiology 143:1321-1326(1997).		
CC	-1- FUNCTION: INVOLVED IN POSITIVE REGULATION OF THE LEVANASE		
CC	OPERON WHICH COMPRISES THE LEVDEFG GENES FOR A FRUCTOSE PTS		
CC	SYSTEM, AND SACA FOR LEVANASE.		
CC	-1- PTM: PHOSPHORYLATED AND INACTIVATED BY VIA THE PHOSPHOTRANSFERASE		
CC	SYSTEM (POSSIBLE).		
CC	-1- SIMILARITY: THE N-TERMINAL REGION CONTAINS A SIGMA-54 FACTOR		
CC	INTERACTION ATP-BINDING DOMAIN		
CC	-1- SIMILARITY: IN THE C-TERMINAL HALF, BELONGS TO THE TRANSCRIPTIONAL		
CC	ANTITERMINATOR BGGL FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation		

CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (See http://www.ebi.ac.uk/join/) or send an email to license@ebi.ac.uk .
CC		-----
DR	EMBL:	M60105; AAA22572.1; -
DR	EMBL:	X92868; CA63460.1; -
DR	EMBL:	Z99117; CAB14650.1; -
DR	PIR:	A3160; A39160.
DR	Sublist:	BGI0677; levr
DR	InterPro:	IPIR001550; -
DR	InterPro:	IPIR002078; -
DR	Pfam:	PF00874; Bgic_antitermin_1.
DR	Pfam:	PF00158; SigmA54_1.
DR	PROSITE:	PS00654; ANTITERMINATORS_BGIG_1.
DR	PROSITE:	PS00675; SIGMA54_INTERACT_1; FALSE_NEG.
DR	PROSITE:	PS00676; SIGMA54_INTERACT_2; 1.
DR	PROSITE:	PS00688; SIGMA54_INTERACT_3; FALSE_NEG.
DR	PROSITE:	PS50045; SIGMA54_INTERACT_4; 1.
KW	Transcription regulation;	Activator; DNA-binding; ATP-binding;
KW	Phosphorylation.	-
FT	DOMAIN	120 .. 351
FT	NP_BIND	148 .. 155
FT	NP_BIND	218 .. 227
FT	SIGNAL	411 .. 572
SO	SEQUENCE	938 AA: 106186 MW: 167025524E7B5696 CRC64: TO THE N-TERMIN OF BGIG FAMILY.

Query Match 3.98; Score 112; DB 1; Length 936;
Best Local Similarity 21.08; Pred. No. 0.39;
Matches 130; Conservative 100; Mismatches 213; Indels 176; Gaps 46.

```

OY      3  SNTMLQRTLLILLSSVYTWMMIFLISQNTATYLSALNLST--SVHWNNKSAKSLIFPKTSLI  61
Db      53  NNLVSKSKVKIKITPPVRYIPVELAEKLENNKWOTEMMEKCOLAFQSNQKQHJISHTN  112
OY      62  PLKPLFTELELRKEIEERKLDQOI--PPRPFTHWNTTSAHSATILNLEFVTCALDIL  119
Db      113  PLFLMIAQKSLIKKAIISOAKAAVFPP-----NGHH--MLLIGPTGS---QKSL-  156
OY      120  ILLEPRHLDQRQKGGDFLRARSSPALTAGACGKMPFNQGTLY-----SFT  169
Db      157  ----FANKITQFATY--SDILKA--GAFLPTTNC---DYNNPQLILSLPLQHKKNSFT  207
OY      170  LPEWQGVSL-----SLDL---IH--PSEGAALMARBARQ--YKILFEKQEPN--TSHV  219
Db      206  GAEDKAGLVEQANGILFMEDEIHLRPPQEGMLFYIISQSYNL---GESHKRTSNV  262
OY      217  FLEGSLTSLNMLECEYIDRDDEAFYCKKQ--HMP-----CELLYMTTARE  264
Db      263  LFICATENSSALL-----KTLRRIPTHTIHPSEHSLSNERVDTLITFLCKEAE  314
OY      265  ----VSYLTDKENSLEPHRSKVG--VEMKMDR-----KHLDVNCNKKRKEILETCGVAKMF  318
Db      315  RIKKSLSVNHIADYVNNALHSAFAFGNVOLKSNVUOLVCAHGFNLNDRNEVELTIVR----  369
OY      314  RYPGAGYLQCKMTTFECNQ-----VGLDTKINQCKLQKLLYLLGD  464
Db      370  DLDPD--EIKEMEMSSKKMORSAISEYVNTTILSIPIYEDDTKITEFDY--STNPHLLE  427
OY      355  STLKRWIYIEPKVAVTLKFDOLHEGTIEKK-----HLLDAEPHTQLCKKHSYPPV79L  416
Db      428  K-----VKTLM-----KEGLSKKDINOJILYTVLHJHVSFF--HHAQVCEKNL  468
OY      411  YSLIDHDYIPREILDSCKKATVITFQO---HPRPPIDIFRRAIGVQAILEFLR  467
Db      469  LPEVDEDTV--QMTKOLKEIAHELDQDTRKRFYIFLSMILGFLKRG---KQIVLVNIG  524
OY      468  SPATKVIKTEINREMIETIER---FGD--PHCY-----IHLIIM-----KQIFK  508
Db      524  -----ETDELRDTHVKEEVRAMIFDKIOEFVFAVLEIEVYIYIMLHNSKSKEN  575

```

QY 509 LAMGILAMDMITAVCTDT 527
 DB 576 KRVLITVA-----AHGNS 589

RESULT 3

YQ2L_YEAST STANDARD; PRT: 1237 AA.

ID YQ2L_YEAST

AC P51254:

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-MAY-2000 (Rel. 39, Last annotation update)

DE HYPOTHETICAL 140.5 KDA PROTEIN IN CTT1-PRP31 INTERGENIC REGION.

GN YGRO90W.

OS Saccharomyces cerevisiae (Baker's Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;

RN 111

RP SEQUENCE FROM N.A.

RA Medler H., Scharfe M., Medler E., Mambuli R.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

RN 121

RP SEQUENCE FROM N.A.

RA Hernandez K., Weber N., Wipfli P., Schmidheini T.;

KL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

CC -1 SIMILARITY: TO S. POMBE SPCC776.08C.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb.slb.ch/announce/>

CC or send an email to license@slb.slb.ch).

CC

DR EMBL: 272975; CAA97093.1;

DR SCD: S0003322; YGRO90W.

KM Hypothetical protein.

SEQUENCE 1237 AA: 140484 MW: 9A285C885493D7D3 CRC64:

Query Match 3.6%; Score 103.5; DB 1: Length 1237;

Best Local Similarity 19.3%; Pred. No. 2.8; 175; Indels 165; Gaps 28;

Matches 102; Conservative 87; Mismatches 15; Gaps 28;

QY

30 NFKLMSALNLSIVHWNNSAKSLFPKTSILPLKPLTETELRIKEITEKLDQOIPRP 89

DB 484 NEYKTRFSINLLIOFPYKVEPKKLLPNRNCI-----RIHQ--ESKEQSLPATPL 331

QY 90 THVNTTTSATSTATILNPRDTYCGDQDLLEVRDHGQRKQYGGDFLRAM----- 143

DB 332 YNFSVLSSTH-----ENLKR-----YLYKTRKQTESFVATYLGRLM 370

QY 144 -----SSPALTAGASCKVMDFNNGTYLVSFTLFWF-----GVQSLLLIPSEGASALM 193

DB 371 QVRGFSNMHSSSLG-----GGTFF--EFTLMAALLNNGGINSKLLH----- 414

QY 194 RARNOGYDKT--IFKG--KRYNGTSNHFTECGTLTNSAAE-----LCGYLDDROGFAYC 244

DB 415 -----GFSYVLEFGVAKYI--ATMDLCHDGHLOPHSNKPESSSSPASKYIDEGQPTTLE 468

QY 245 MKRQHPCEALTYMTTHNREV--SYLTDKENSLEFHSKSVQVEMKDR--KHIDVYCNCKRE 301

DB 459 DKSTRV--NLTKMTVSSYOILKEVAGETLMLNN-----VVODQSFNIFLTNISRD 519

QY 302 KLEHCUVGKMPRPVGGYTLOKMKITFCNOVOYDTIKINGLCKKLLYLIGDSTLRQW 361

DB 520 NLK-----YDL-----CYDVPPLQCKYNN--LESLATATGSMERYKFI 556

QY 362 YIFPRVAVTLKFDLHETGIFKKHLLDAERHIOIOM--KKHSYFPVTQOLYS-----LI 414

DB 557 -----TLNPLAKRITNVARVALGDRIRYIOIEWVGOKSDPITRKRYSYNTGDNH 608

QY 415 DHDYI-----PREIDRL--SGDKNAIVITGQHFRRPPIIDIFIRRAQVOKAIE-- 463

DB 609 NEDFVAVLLIYNPSECDKLVTKGPAHSETMSTAFAVKNF-----MGIKSSLRPK 659

QY 464 -----LFLRSPATVVIKTENI-----RPMHIERPRGDHGY 497

DB 660 DGSITHCVMSTSSSEPIITSSIVNALQKHVSKKAQISNLTIKKHNFL 708

RESULT 4

COX1_TETPY STANDARD; PRT: 698 AA.

ID COX1_TETPY

AC P11947;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).

GN COI.

OS Tetrahymena pyriformis.

OC Mitochondrion.

OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;

OC Tetrahymena; Tetrahymena.

OX NCBI_TaxID=5908;

RN 111

RP SEQUENCE FROM N.A.

RC STRAIN-ST;

RX MEDLINE-88184706; PubMed-2833363;

RA Zlate Z., Suyama Y.;

RT "The cytochrome oxidase subunit I gene of Tetrahymena: a 57 amino acid NH₂-terminal extension and a 108 amino acid insert."

RL Curr. Genet. 12:357-368(1987).

-1 FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY

CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1

3 FORM THE FUNCTIONAL CORE OF THE ENZYME. ELECTRONIC ORIGINATING IN

CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONIC ORIGINATING IN

CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2

AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3

AND COPPER B.

-1 CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +

4 FERROCYTOCHROME C.

CC

-1 PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.

-1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL

INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.

-1 SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb.slb.ch/announce/>

CC or send an email to license@slb.slb.ch).

CC

EMBL: L28677; AAA32102.2;

DR EMBL: X06133; -; NOT_ANNOTATED_CDS.

DR PIR: S00742; S00742.

DR HSP: P00396; 10CC.

DR InterPro: IPR000883; -.

DR Pfam: PF00115; COX1; 2.

DR PROSITE: PS00077; COX1; 1.

KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;

KW Respiratory chain; Inner membrane.

FT METAL 111 111 IRON (HEME A) (PROBABLE).

FT METAL 401 401 COPPER B (PROBABLE).

FT METAL 405 405 COPPER B (PROBABLE).

FT METAL 450 450 COPPER B (PROBABLE).

FT METAL 451 451 COPPER B (PROBABLE).

FT METAL 536 536 IRON (HEME A3) (PROBABLE).

FT METAL 538 538 IRON (HEME A) (PROBABLE).

FT METAL 698 698 IRON (HEME A) (PROBABLE).

SEQUENCE 698 AA: 81746 MW: D6B0F48BA3109A72 CRC64:

Query Match 3.4%; Score 100; DB 1; Length 698;
 Best Local Similarity 19.2%; Pred. No. 2.4;
 Matches 105; Conservative 81; Mismatches 194; Indels 166; Gaps 30;

OY 23 MIFLISQNTKLM-----SALNLSISVH---YMNNSAKSLFKETSLIAKLETELRIK 74
 DB 1 MLEIFV-SFDMWVDFEIOKSPKSVNNNYIYNKIKLF--TYLNDLR-----K 48
 OY 75 EIEKLDQOIPPRF-----THVNTTSATSTATTILNPDIYCRQDQDILLEY 124
 DB 49 HILKTYVTIHNKRIAINLYLSMVTGLSCALATWIRBELAHPESPFPGDSLRLQVY 108
 OY 125 RDLGQRKQYOGDFLRARMSPALTAGASGVDPNNGIYLSFTL-----FWE---GVV 176
 DB 109 TAH-----GLIMVFVYVPILEGGFANFLIPYVSGKDAIPRLNIGFIOPCYI 160
 OY 177 SLSTLL-LIHPSGASALHARNOGYDKIIFKGFVNGTSHVTECGGLTNSNAELCEYLD 235
 DB 161 LIAKIGFLRPQ-----FWRV-----YDKTSFSFPFLKKMY-----NQYKEKN 199
 OY 236 D-----RDOEAFY-----CMKPOH-----MPCALYMTNRREVSYLTKE 272
 DB 200 DYLEFLDLKKEITDDHSFPMARKVYKLPQYVSFPFLKMMKMTINPESP-----254
 OY 273 NSLFHRSKVGVEKMKRKHIDYTCNKRKIEBTOCGVAKPPVGGYTLQGMITTFGNO 332
 DB 255 --WYASRV---VOSRRKKVFTKCSARL-----TTAGWT-----FITPFS- 292
 OY 333 VQDITKINGCLKGLIYLLDSTLRQMIYFPKVVYTKLFKEGTFKKHLLDER 392
 DB 293 ----NIKYTG-----VGSQDILLSVFAGSTISFTNL---ITRNTLMPQMR 336
 OY 393 HTQIQMKKSHYPTVTFOLYSLIDHDYIPREIDRLSGDKNTAIY-ITFGGHEPPEPIDFI 451
 DB 337 HRAV-----LMPTVITISIFTL-----MLATITPVGAIVTMAFDRHWQ---TTF 381
 OY 452 RRAIGVOKAIEH---LFLSPATKVIITKE---NIREMLETERRGDHF-----G 495
 DB 382 EYAGGDPILSQHLFWFGHPEYVLIIPFGFIMIVPNNNRRARASHHMAIYMA 441
 OY 496 YIHLYI 501
 DB 442 YMGYLV 447
 RESULT 5
 Y137_CAEEI STANDARD; PRT; 400 AA.
 ID Y137_CAEEI
 AC P34428;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE HYPOHETICAL 45.5 KDA PROTEIN P44B9.7 IN CHROMOSOME III.
 GN P44B9.7
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida; Rhabdillidae;
 CC Rhabdillidae; Pelodertinae; Caenorhabdilis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister L., Jier M.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopia A., Saunders D., Showkhen R.,
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sultson J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Watson R., Watson A., Welstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/ambrosity/>
 CC or send an email to license@isb-sdb.ch).
 DR EMBL: I23648; AAA28029.1;
 DR PIR: S44810; S44810.
 DR Wormpep; P44B9.7; CE00176.
 KM Hypothetical protein.
 SQ SEQUENCE 400 AA; 45523 MW; 96BDF6C3B4C2D7D CKE64;

Query Match 3.4%; Score 99.5; DB 1; Length 400;
 Best Local Similarity 22.0%; Pred. No. 1.2;
 Matches 70; Conservative 36; Mismatches 129; Indels 81; Gaps 15;

OY 56 PKTSLT-----PLKPLTELEIKELIEKLDQOIPRPFTVNTTTSATISTAT 105
 DB 62 PTLTSLAQLHLSAPLPAPPOOOSHL-----SQLDQPTTSTAINPHAPVOTI 111
 OY 106 LNPDIYCRQDQDILLEYRDHGGORQYOGDFLRARMSPALTAGASGVDPNNGIY 145
 DB 112 ARRDENARNAGLEBELAARQ--QOGEF--MLDQHQQOQNRBAQQLAEVAFAN---142
 OY 166 VSTFLWEQOVSLSLLIHPSGASALHARNOGYDKIIFKGFVNGTSHV---FTGCG 221
 DB 163 -----PPTVSIPTQGHPPQOTSASSQORAOE-----TMAQOTEPUSHIRESQNETN 211
 OY 222 LTNSVAELC-----EYLDPRDOEAFPC-----MKPOHM-----PCEALTYM 278
 DB 211 LEKSYGELCYGRELVDNLVNTSYLSTILKYMKEKPIVNOENPNDAQCCNNAQMR 270
 OY 259 TTRNREVSITDKENSLFHSKYG---VENMKD---RKHIDYTCNKRKIEB--TQCV 309
 DB 271 S-----DIQIIEKREPTWKRMTGDIETLMLDSEILKFMDEKRRALFEOHAIIVI 326
 OY 310 GMPKPPVGGYTLQGMKIT 327
 DB 327 SKTPVEGHIMYGKMTS 344

RESULT 6
 SPAT_BACSU STANDARD; PRT; 614 AA.
 ID SPAT_BACSU
 AC P93116;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE SUBUNITIN TRANSPORT ATP-BINDING PROTEIN SPAT.
 GN SPAT OR SPAB OR SPAY.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 6633;
 RX MEDLINE=92171481; PubMed=1539969;
 RA Klein C., Kaletta C., Schnell N., Entian K.-D.;
 RT "Analysis of genes involved in biosynthesis of the antibiotic
 RT subtilin.";
 RL Appl. Environ. Microbiol. 58:132-142(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 6633;
 RX MEDLINE=92138640; PubMed=1735728;

```

RA  Chung Y.J., Steen M.T., Hansen J.N.;
RT  "The subtilin gene of Bacillus subtilis ATCC 6633 is encoded in an
RT  operon that contains a homolog of the hemolysin B transport
RT  protein."
RL  J. Bacteriol. 174:1417-1422(1992).
CC  -1- FUNCTION: PROBABLY IMPLICATED IN THE EXPORT PROCESS OF THE
CC  LANTIBIOTIC SUBTILIN.
CC  -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC  -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC  (ABC TRANSPORTERS), HLYB SUBFAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: M86869; AAA22838.1;
DR  EMBL: M83944; AAA22770.1; ALT_INIT.
DR  EMBL: M9263; AAA22776.1; ALT_INIT.
DR  EMBL: U09819; AAB91587.1;
DR  PIR: B43935; B43935.
DR  HSSP: P13569; 1NRD.
DR  InterPro: IPR001617;
DR  Pfam: PF00005; ABC_tran; 1.
DR  PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW  Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW  ATP-binding.
FT  NP_BIND 387 394  ATP (POTENTIAL).
FT  CONFLICT 39 39  R -> A (IN REF. 2).
FT  CONFLICT 68 68  E -> D (IN REF. 2).
FT  CONFLICT 394 394  K -> S (IN REF. 2).
FT  CONFLICT 407 408  HE -> QC (IN REF. 2).
SQ  SEQUENCE 614 AA: 71188 MW: 8A2DE2EAD036C33F CRC64:
Query Match 3.48; Score 99.5; DB 1; Length 614;
Best Local Similarity 20.44; Pred. No. 2.2;
Matches 128; Conservative 88; Mismatches 212; Indels 199; Gaps 29;

```

```

QY 378 ETGIFKHLDAER-----HTQI-----QMKHSYPTVPOYS 412
DB 458 QTN--RMHEVDIYVADFLKSHSSYQFDGLGFWDEGRQSGGQMOKIALARAYFEAS 515
QY 413 LIDHDIPEIDRLSGKNATVITFGQH--RPFIDIPRAIGVOKATEIRFLRSPA 470
DB 516 LYIIDPSSALDPAKEK-----TFDFFSLSKDKIGITISHRLVAKLADRIYVDKG 569
QY 471 TKV-----IKTENIREMHIETEPF 490
DB 570 EIVGIGTHELLKTCPEYKKRDESENV 596

RESULT 7
SCAA_RAT STANDARD; PRT: 698 AA.
AC P37089; Q64593;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AMILORIDE-SENSITIVE SODIUM CHANNEL ALPHA-SUBUNIT (EPITHELIAL NA+
DE CHANNEL, ALPHA SUBUNIT) (ALPHA ENAC) (NONVOLTAGE-GATED SODIUM CHANNEL
DE 1 ALPHA SUBUNIT) (SCNEN) (ALPHA NACH).
GN SCNNIA OR RENAC.
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxid=10116;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR; TISSUE-Distal colon;
RX MEDLINE-93170495; PubMed-8382122;
RA Lingueglia E., Volley N., Waldmann R., Lazdunski M., Barbry P.;
RT "Expression cloning of an epithelial amiloride-sensitive Na+ channel.
RT A new channel type with homologues to Caenorhabditis elegans
RT degenerins."
RL FEBS Lett. 318:95-99(1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Colon epithelium;
RX MEDLINE-93156815; PubMed-8381523;
RA Canessa C.M., Horisberger J.D., Rossier B.C.;
RT "Epithelial sodium channel related to proteins involved in
RT neurodegeneration."
RL Nature 361:467-470(1993).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR-KYOTO, AND SHRSP; TISSUE-Kidney;
RX MEDLINE-97191134; PubMed-9039092;
RA Kreitz R., Struk B., Rubattu S., Hubner N., Szpieler J., Szpieler C.,
RA Ganem D., Lindpaintner K.;
RT "Role of the alpha-, beta-, and gamma-subunits of epithelial sodium
RT channel in a model of polygenic hypertension."
RL Hypertension 29:131-136(1997).
[4]
RP TOPOLOGY, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-95014183; PubMed-7929098;
RA Snyder P.M., McDonald F.J., Stokes J.B., Welsh M.J.;
RT "Membrane topology of the amiloride-sensitive epithelial sodium
RT channel."
RL J. Biol. Chem. 269:24379-24383(1994).
[5]
RP TOPOLOGY.
RX MEDLINE-94230383; PubMed-8175716;
RA Renaud S., Lingueglia E., Volley N., Lazdunski M., Barbry P.;
RT "Biochemical analysis of the membrane topology of the amiloride-
RT sensitive Na+ channel."
RL J. Biol. Chem. 269:12981-12986(1994).
[6]
RP MUTAGENESIS OF SER-569 AND SER-593.
RX MEDLINE-95263507; PubMed-7744818;
RA Waldmann R., Champigny G., Lazdunski M.;

```

```

RT      *functional degeneratein-containing chimeras identify residues essential
RT      for amiloride-sensitive Na+ channel function.*;
RL      J. Biol. Chem. 270:11735-11737(1995).
CC      -1- FUNCTION: SODIUM PERMEABLE NON-VOLTAGE-SENSITIVE ION CHANNEL
CC      INHIBITED BY THE DIURETIC AMILORIDE. MEDIATES THE ELECTRODIFFUSION
CC      OF THE LUMINAL SODIUM (AND WATER, WHICH FOLLOWS OSMOTICALLY)
CC      THROUGH THE APICAL MEMBRANE OF EPITHELIAL CELLS. CONTROLS THE
CC      REABSORPTION OF SODIUM IN KIDNEY, COLON, LUNG AND SWEAT GLANDS.
CC      ALSO PLAYS A ROLE IN TASTE PERCEPTION.
CC      -1- SUBUNIT: HETEROTETRAMER OF TWO ALPHA, ONE BETA AND ONE GAMMA
CC      SUBUNIT. A DELTA SUBUNIT CAN REPLACE THE ALPHA SUBUNIT.
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN
CC      -1- SIMILARITY: BELONGS TO THE AMILORIDE-SENSITIVE SODIUM CHANNELS
CC      FAMILY
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X70521; CAA49916.1; ALT_INIT.
DR      EMBL; X70497; CAA49905.1; -.
DR      EMBL; U54699; AAB61156.1; -.
DR      EMBL; U54700; AAB61157.1; -.
DR      PIR; S29499; S29499.
DR      InterPro: IPR001873; -.
DR      Pfam: PF00858; ASC. 1.
DR      PRINTS: PR01078; AMINACHANNEL.
DR      PROSITE: PS01206; ASC. 1.
KM      Ionic channel; Transmembrane; Ion transport; Glycoprotein.
FT      DOMAIN 1 110
FT      TRANSMEM 1 131
FT      DOMAIN 132 589
FT      TRANSMEM 590 610
FT      DOMAIN 611 698
FT      CARBOHYD 190 190
FT      CARBOHYD 259 259
FT      CARBOHYD 320 320
FT      CARBOHYD 339 339
FT      CARBOHYD 424 424
FT      CARBOHYD 538 538
FT      MUTAGEN 588 588
FT      MUTAGEN 592 592
FT      CONFLICT 598 599
FT      SEQUENCE 698 AA; 78887 MW; BOCF7C15G3CE9763 CRC64;

```

Query Match 3.44; Score 98; DB 1; Length 698;

Best Local Similarity 19.24; Pred. No. 3.5; Indels 164; Gaps 18;

Matches 89; Conservative 53; Mismatches 157; Indels 164; Gaps 18;

```

QY      11 LLIISFVYTWMIPIISQNTKLMGALNLSIVHYWNNSAKSLPKFSLPLPLTE 70
DB      114 VLMCTFGMMYQFALLFEELYSLVPSLNINL-----NSDKLVPAVYVCLNDP 167
QY      71 LRKEIKELKDQ-----QIRPRPF 89
DB      168 --IKEELEELDRIETQTLFDLYKNSYTRKAGARRSSDGLGAPPHLQRLTPPE 225
QY      90 THWTTTSTHSTATILNPR-----DYCRGDQDILLLEVROH 127
DB      226 S--GRTASGSSSVNDNPNVDKDKKIGFOLCONNSDCFYOTYSSG--VDAYRE---- 277
QY      128 LGQRKQYGGDFLRAMSSPALTAGASGKYDFFNNGTIVSFTLFFNGGVSLILLHPSE 187
DB      278 -WYRFHYINILSRISDTSPLALEEALG-----NFTTCFQNPAPCNOANIS-KFHHPK 329

```

```

QY      168 G-----ASLRAPNOGYDKILFKGFVNGSHWFTECGIILNSNAELCEVLDRO 246
DB      330 GNCYFENDKNNKNSLWMSMPG-----VNN-----SLSTLFTEDNDP1----- 347
QY      239 QEAFFYCKMPDQHPHPCALTYMTTRNRESVLTDEKNSLPHRSKGVENMKRKHII VTN 246
DB      368 -----PLSTVIGARVYVHGQDEPARFMDGCGNLPVGVENSISMKKEA--LISLQGN 417
QY      299 KREKIEETQYCKMKRPVPGSYTLLOGKMTTPCNQVOLDI KINSGCKKLLYLDLSTLE 346
DB      418 YGDCTEGSDVPEKKNLYPSKYTQD-----VCISCFCEENMKKCCGA--- 454
QY      359 QWITY-FPVVYKTLKF-----FDLEHETGIFKK 384
DB      460 -YIFPVPKGVFECDFYKROSSWGYCYKLGCAFSLDSIGCFSK 501

```

RESULT 8

```

HUI4_YEAST
ID      HUI4_YEAST          STANDARD;          PRT;          892 AA.
AC      P40985;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      PROBABLE UBIQUITIN--PROTEIN LIGASE HUI4 (Ec 6.3.2.-).
GN      HUI4 OR YJR036C OR J1608.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=5288C;
RX      MEDLINE=95397595; PubMed=7668047;
RA      Huang M.-E., Chast J.-C., Galibert F.;
RT      "Analysis of a 42.5 kb DNA sequence of chromosome X reveals three
RT      tRNA genes and 14 new open reading frames including a gene most
RT      probably belonging to the family of ubiquitin-protein ligases."
RL      Yeast 11:775-781(1995).
RN      [2]
RP      SEQUENCE OF 362-892 FROM N.A.
RX      MEDLINE=95045431; PubMed=7957102;
RA      van Gool A.J., Verhage R., Swaenmakers S.M.A., van de Putte P.,
RT      Browner J., Troelstra C., Bootsma D., Hoeijmakers J.H.J.;
RT      "RAD26, the functional S. cerevisiae homolog of the Cockayne syndrome
RT      B gene ERCC6."
RL      EMBO J. 13:5361-5369(1994).
RN      [3]
RP      GENE NAME, AND GENE DISRUPTION.
RX      MEDLINE=9607972; PubMed=9858558;
RA      Wang G., Yang J., Hultberg J.M.;
RT      "Functional domains of the rps ubiquitin-protein ligase."
RL      Mol. Cell. Biol. 19:342-352(1999).
RN      [4]
RP      FUNCTION: PROBABLE E3 UBIQUITIN-PROTEIN LIGASE WHICH ANCHORS
RT      UBIQUITIN FROM AN E2 UBIQUITIN-CONJUGATING ENZYME IN THE FORM OF A
RT      THIOESTER AND THEN DIRECTLY TRANSFERS THE UBIQUITIN TO TARGETED
RT      SUBSTRATES. NON ESSENTIAL.
CC      -1- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR
CC      UBIQUITIN-THIOESTER FORMATION (BY SIMILARITY).
CC      -1- SIMILARITY: CONTAINS 1 HECT-TYPE E3 UBIQUITIN-PROTEIN LIGASE
CC      DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L76344; AAB8738.1; -.
DR      EMBL; Z49356; CAA89563.1; -.

```



```

RESULT 10
EX5C_HAEIN STANDARD; PRT: 1121 AA.
ID EX5C_HAEIN
AC P44945;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE EXODOXYRIBONUCLEASE V GAMMA CHAIN (EC 3.1.11.5).
OS RECC OR H10942
OC Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KM20 / ATCC 51907.
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Keriavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Seudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Furman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: EXHIBITS A WIDE VARIETY OF CATALYTIC ACTIVITIES
CC INCLUDING ATP-DEPENDENT EXONUCLEASE, ATP-STIMULATED ENDONUCLEASE,
CC ATP-DEPENDENT HELICASE AND DNA-DEPENDENT ATPASE ACTIVITIES.
CC (BY SIMILARITY)
CC -1- SUBUNIT: CONSIST OF THREE SUBUNITS; RECC, RECC AND RECC
CC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32775; AAC22596.1; -
DR TIGR: H10942; -
KM Hydrolyase; Nuclease; Exonuclease; Endonuclease; Helicase;
KM DNA repair.
SQ SEQUENCE 1121 AA; 129668 MW; E5070957296AED03 CRC64;

Query Match 3.48; Score 98; DB 1; Length 1121;
Best Local Similarity 18.98; Pred. No. 6.8;
Matches 102; Conservative 97; Mismatches 204; Indels 136; Gaps 31;

OY 49 NSAKSLFPTSLIPKLPTELRIKEIIEKLDQIPREFHVNTTSAT---HSTAF 104
DB 411 NODPSLTPKDVVAVADINQYPIQAVFGKNDVQPIFSLSNKTSSEDLVSSYLT 470
OY 105 ILNPRDYCRGDQDILLVYRDLGQRQYQ---GD--FLARMSPPALTAGASGV--M 157
DB 471 LFLAKSNFSAEDVLVLDLP---AMREFNISLADLPVREMYVDSGIRFGLOKQNDGI 527
OY 158 DFNNGTYLVSTFLFMEGQVSLILLIHPSEGASALMR---ARNQGYOKILFKKFNQGS 214
DB 528 NFNS-----WQAGLE-RNIIIGYAMREGEINQDGLGINSYG---LKEGLAGNIS 573
OY 215 HVTEGGLTINSNALECEYLDROEAFYCKRQHPHCALITYMTNRREVS---YLTD 270
DB 574 HFTALSA-----LHETLQQAHSIEKQWELTALLSDFFVNNEDTSDMIRYIOE 622
OY 271 KENSLFHSKSYGVEMKDKRHIDVTNCKNR---EKIEETQGVGMKPPVPGGYTLQCKWIT 327

```

```

DB 623 KINEL-----AEHLKTLHNEELQAEVIAVITWQLEADPNSLKPACK--V 667
OY 328 TFCNOVOLDITIKINGCLKKGLIYLDGSDTRQWIIYFEPVYVATKTFDELHETIHK--- 383
DB 668 NCCITLPMRSVPE-----KVYCLGNNDAD-----YPR-VUTRNSLIMQHYKGNIBV 719
OY 384 -----KHLILDARHTQIQOMKKHSFPVPTOLYSLIDHDTYIRE-----ILRUS-GD 429
DB 716 RDDDRITFLER-----LLAARDYCIYSVGRSITDNO--PKRPSVLSQLLIY INQGO 767
OY 430 KNTAIVTFEQH-FRFPIDIFIRRAIGVQKATERFLKSPAIK----- 472
DB 768 KEN--VLTVIEHPMTAFSPDNF-----KNNKK-FTRSPATWLPILAFVASSNSSEF 816
OY 473 VITENIREM-HIEERFGGF-HGYIHTLTKDIT---FKQLNGLITAMENTIAVYIL 926
DB 817 AVTMTLEKIEVELDALVSFVENPVKFEFKQGVYFKOKERIASDSNFTLS-GID 874

RESULT 11
ID NPH3_HUMAN STANDARD; PRT: 221 AA.
AC 095157; Q9ULR1;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUREXOPHILIN 3 (FRAGMENT).
GN NXP3 OR NPH3 OR KIA1159.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039618; PubMed=10574461;
RA Hirosewa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,
RA Ohara O.;
RT "Characterization of cDNA clones selected by the GeneMark analysis
RT from size-fractionated cDNA libraries from human brain."
RL DNA Res 6:3329-336(1999).
RN [2]
RP SEQUENCE OF 49-221 FROM N.A.
RX PubMed=9570794;
RA Müssier M., Suedhof T.C.;
RT "Neurexophilins form a conserved family of neuropptide-like
RT glycoproteins."
RL J. Mol. Neurosci. 18:3630-3638(1998).
CC -1- FUNCTION: MAY BE SIGNALING MOLECULES THAT RESEMBLE NEUROPEPTIDES
CC AND THAT ACT BY BINDING TO ALPHA-NEUREXINS AND POSSIBLY OTHER
CC RECEPTORS (POTENTIAL).
CC -1- SURCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -1- TISSUE SPECIFICITY: BRAIN-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO THE NEUREXOPHILIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB012985; BAA86473.1; -
DR EMBL: AF043468; AAD02281.1; -
DR MIM: 604636; -
KM Glycoprotein.
KW NON-TER
FT CARBOHYD 31 1
FT CARBOHYD 96 96
FT CARBOHYD 106 106
FT CARBOHYD 112 112

```


MSH6_ARATH
 ID MSH6_ARATH STANDARD: PRT: 1324 AA.
 AC 004716;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE DNA MISMATCH REPAIR PROTEIN MSH6-1 (ATMSH6-1).
 GN MSH6-1 OR AGAA.3 OR AT4G02070 OR T10M13.8.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 CC Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_TaxID=3702;
 RN 1;
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, LANDSBERG ERCTA;
 RA Till S., Granat S., Parnell L., Kaplan N., Hoffman J., Lodhi M.,
 RA Johnson A.F., Dedhia N., Martienssen R., McCombie W.R.;
 RL Submitted (May-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Ade J., Belzile F., Philippe H., Douriaux M.P.;
 RT *Four mismatch repair paralogs coexist in Arabidopsis thaliana:
 RT ATMSH2, ATMSH3, ATMSH6-1 and ATMSH6-2.*;
 RL Mol. Gen. Genet. 262:239-249(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Ertlan K.-D., Terry N.,
 RA Harris B., Ansoorge W., Brandt P., Griwell L., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
 RA Kreis M., Dalseny M., Puigdomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohenseil U., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bihlan L., Robben J.,
 RA Van der Schueren J., Glymonprez B., Chang Y.-J., Vandenbussche F.,
 RA Braeken M., Welljens I., Voet M., Baetliens I., Aert R., Defoer E.,
 RA Wiltzenegger T., Bothe G., Rampsperger U., Hilbert H., Bruun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirse W.,
 RA Mooijman P., Klein Lankhorst R., Rose M., Haut J., Koetler P.,
 RA Bernerster S., Hempel S., Felpusch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIay K., Mayes R.,
 RA Borkova D., Rajandream M.-A., Lyne M., Benes V., Reichtmann S.,
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartman B., Granderath K., Danner D., Herzi A.,
 RA Neumann S., Argilou A., Vitale D., Liguori R., Pauet E., Herzi A.,
 RA Massenet O., Ougleir F., Clabaud G., Mündlein A., Felder P.,
 RA Schödel S., Hiller R., Schmidt W., Lecharry A., Aubourg S.,
 RA Giddons T., Weber N., Berger C., Montfort A., Casacuberta E.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bietke C.,
 RA Frisman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Hadelmann K.,
 RA Parnell L., Dedhia N., Gof L., Schutz K., Huang E., Spiegel L.,
 RA Sehkon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Clond J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Giesel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drome K., Cotton M., Joshi C.,
 RA Antonola B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Yil D., Stekter M., Matero A., Shah R.,
 RA Svedy I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Grant S., Shindy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,

RA Chen E., Marra M., Martienssen R., McCombie W.R.;
 RT *Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RU thaliana.*;
 RU Nature 402:769-777(1999).
 CC -1- FUNCTION: INVOLVED IN POST-REPLICATIVE DNA-MISMATCH REPAIR (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: HETERO-DIMER OF MSH2 AND MSH6 (OTHER) (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MOTIF FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/about/cv/>
 CC or send an email to license@isb.ch).
 CC
 DR EMBL: AF001535; AB57796.1;
 DR EMBL: AJ245967; CAB5337.1;
 DR EMBL: AF001308; AAC78699.1;
 DR EMBL: AL161493; CAB80700.1;
 DR InterPro: IPR000432;
 DR InterPro: IPR002863;
 DR Pfam: PF00488; Muts_C; 1.
 DR Pfam: PF01624; Muts_N; 1.
 DR PROSITE: PS00486; DNA_MISMATCH_REPAIR_2; 1.
 DR DNA repair; ATP-binding; DNA-binding.
 KW NP_BIND; 1083
 FT NP_BIND 316 318 ATP (POTENTIAL).
 FT CONFLICT 383 383 Q -> OVERAAGN (IN REF. 1).
 FT CONFLICT 852 852 P -> PGILNLIHILILIANCTASHIISLP (IN
 FT REF. 1).
 SQ SEQUENCE 1324 AA; 146797 MW; 7874340272C9796 CRC64;
 Query Match 3.34; Score 97; DB 1; Length 1324;
 Best Local Similarity 21.24; Pred. No. 10;
 Matches 84; Conservative 68; Mismatches 126; Indels 318; Gaps 24;
 QY 123 EVRDHLGQKRGYGG--DELRARMSFALTAG--ASGKYMPFN-NGTYVSFTLPEKSVS 177
 DB 460 ETPOLEORRKRKTSKDKVYREVCVAVTKGTLTGDGELLINDAYILALTI---EGES 516
 QY 178 LSLILHPESGASALMRARNGYDKIIFRGKVNQTSIVTEBGLT-NSNNELEHLD 246
 DB 517 LT-----NPT-----ASHNFGVNVAVATUKILDKEDK 546
 QY 237 RDEAFYCC--MKRQHM--PCEALTYMT-----TRNEVSYLTI-----TPPCNV 433
 DB 547 QDCSALSCLSLSEMRVELIKPAKVLVATERIVQTRNPLNVNLTPLSHWDSEKTIY- 605
 QY 278 RSVGVYEMKDKRHIDVTCNKRREKIEETCGVKKPPVPGGYTLQCKNI---TPPCNV 433
 DB 606 -EVGIIYR-----INCP-----SSAVSECKILGDSSEFIFKML 430
 QY 334 -QIDTIKINGCLK---GKLTLYLGDSTLKQMIYFPKVVYLIKFTLHETGIFKHLIL 348
 DB 641 SETATDKKSGSLAALGAIYLRAPAFDSLRLFAK-FESIPACDSNVN-EKQHMVL 698
 QY 389 DAEKRTQIQMKHS---YFVFYFQYSLIDH-----DVIPHEI--DLSQDN 441
 DB 699 DAAALENLEIFERNNGY---SGLTVLQNOGCTASGRKLKLTWLPRLINTELKEW 755
 QY 432 TAYVITFGQHRPPPIIDIFRRARIGVQAIERFLR 467
 DB 756 DVAALIRGEN---LPSLEFRKSLSLPDMERLIAR 788
 RESULT 15
 P2X2_HUMAN
 ID P2X2_HUMAN STANDARD: PRT: 471 AA
 AC G09B09; 09Y637; 09Y638; 09Y639; 09Y640; 09Y641; 09Y642; 09Y643;
 DT 01-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Ref. 40, last sequence update)
 DT 01-OCT-2000 (Ref. 40, last annotation update)
 DE P2X PURINORECEPTOR 2 (P2X2) (PURINERGIC RECEPTOR).
 UN 1.2RX2 OR P2X2.
 UN Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCRL_TaxID=9605;
 RN [1]
 RP SOURCE: FROM N.A. (ISOFORMS A; B; C AND D).
 RC TISSUE-Placental;
 RX PubMed=10570044;
 RA Lynch K.J., Touma E., Niforatos W., Kage K.L., Burgard E.C.,
 RA van Biesen T., Kowalik E.A., Jarvis M.F.;
 RT "Molecular and functional characterization of human P2X2(2)
 RT receptors";
 RL Mol. Pharmacol. 56:1171-1181(1999).
 RN [2]
 RP SOURCE: FROM N.A. (ISOFORMS A; C; H AND I).
 RC TISSUE-Prostate;
 RA Chang T.K., Kosaka A.H., Oglesby I.B., Geyer J.R., Lachnit W.G.,
 RA Ford A.P.D.W., Chang D.J.;
 RT "Cloning and molecular characterization of human P2X2 and its splice
 RT variants";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SOURCE: FROM N.A. (ISOFORMS A AND B).
 RC TISSUE-Placenta;
 RA McMahon R.A., Egan T.M., Hurley P.T., Nelson A., Rogers M., Martin F.;
 RT "Cloning of the human P2X2 receptor cDNA and multiple splice
 RT variants";
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDING OF THIS LIGAND GATED ION CHANNEL TO ATP
 CC MEDIATES SYNAPTIC TRANSMISSION BETWEEN NEURONS AND FROM NEURONS TO
 CC SMOOTH MUSCLE.
 CC -1- SUBUNIT: HOMO- OR HETEROPOLYMERS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS: A (SHOWN HERE), B, C, D, H AND
 CC I; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE P2X RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF190822; AAF19170.1; -;
 DR EMBL: AF190823; AAF19171.1; -;
 DR EMBL: AF190824; AAF19172.1; -;
 DR EMBL: AF190825; AAF19173.1; -;
 DR EMBL: AF190826; AAF19174.1; -;
 DR EMBL: AF260426; AAF74201.1; -;
 DR EMBL: AF260427; AAF74202.1; -;
 DR EMBL: AF260428; AAF74203.1; -;
 DR EMBL: AF260429; AAF74204.1; -;
 DR EMBL: AF109387; AAD42947.1; -;
 DR EMBL: AF109388; AAD42948.1; -;
 DR MIM: 600844; -;
 DR InterPro: IPR001429; -;
 DR InterPro: IPR003045; -;
 DR Pfam: PF00864; P2X_receptor.1;
 DR PRINTS: PRO1307; P2XRECEPTOR.
 DR PROSITE: PRO1212; P2X2RECEPTOR.
 DR PROSITE: PS01212; P2X2RECEPTOR.1;
 KM Ionic channel; Transmembrane; Ion transport; Receptor; Glycoprotein;
 KM Alternative splicing.
 FT DOMAIN 1 42 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 43 63 1 (POTENTIAL).
 FT TRANSMEM 64 317 EXTRACELLULAR, CYSTEINE-RICH (POTENTIAL).
 FT TRANSMEM 318 358 2 (POTENTIAL).

FT DOMAIN 359 471 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 320 333 PORE-FORMING MOTIF (POTENTIAL).
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPPLIC 36 127 MISSING (IN ISOFORM H).
 FT VARSPPLIC 38 152 NRRGLVYRAVOLILLYFYWVFIYOKSYDSESGESSI
 FT ITKKKITTSEHKVWDEVEYKPEEGSYFSTITRVATHS
 FT OTQGTGPERISRYNATCLSDACAGELMUGN -> ITHRA
 FT EKLDEGDEPRELHHQGDHHAQSVAGRGVREAPR
 FT (IN ISOFORM I).
 FT VARSPPLIC 104 127 MISSING (IN ISOFORM C).
 FT VARSPPLIC 354 354 V -> VWRNPLMGPGCGGSTRPLHTGLCPQ (IN
 FT ISOFORM D).
 FT VARSPPLIC 381 447 MISSING (IN ISOFORM B).
 FT CONFLICT 1 14 MAAQPKYAGATA -> MV (IN REF. 3).
 SO SEQUENCE 471 AA; 51754 MW; 84CDB1DA136EF420 CRC64;

Query Match 3.3%; Score 96.5; DB 1; Length 471;
 Best Local Similarity 21.4%; Pred. No. 2.7;
 Matches 88; Conservative 46; Mismatches 120; Indels 157; Gaps 23;

QY 11 LLLLSFSVYWMFIISQNTFKLMSALNISVHWNNASAKSLFPTSLI-PLKPLTFT 69
 DB 50 LLLILYF---VWVFIYOKSYOE-----SETGPESSITIVKGIITS 88
 QY 70 ELRIKEIIEKLDQIPRP-----FTHWNTTSATHT-----ATILNPROTYCR 114
 DB 89 EKHVMDVEEYK---PREGSVFSITTEVHTISQGTGCPESIRVHNATCLSDAD--CV 143
 QY 115 GQDILLEVHDHGRKQYGDPLRA-----KMSPALTAGAS-----KWDFFNN 161
 DB 144 AGELEML-----GNGLRTRCVPYQPSKTCVEFGWCVEDGASVQF-L 188
 QY 162 GTIYVFTLFMEGVSTSLLLIHPEGASALMARNOGYDK-----IIFK-GK 208
 DB 189 GTVAPNFIILKN--SIHYKPFHFKGNIA---DRDGYLKNCTFHEASDLYCPFKLGF 243
 QY 209 FVNGSHVTE-----CGILNLSNAELCYLDDDDQAFRC-----MPCMPPE 253
 DB 244 IYKAGESFTLAHKGGVIGVLIIMDCD-----DLPASCNPKYSFRLDPKHVPAS 296
 QY 254 A-----LTMYTTR-----REVSYLTKENSLFHRSKYGV-E 284
 DB 297 SGVNFRAKYKINGTTRTLKAVGIRIDIVYHQAQKFSILPTIINLATATLSVGVS 356
 QY 285 MKDKRKHIDVNCNK---REKIEFCQ-----VGKKPPVPG 317
 DB 357 FICDMLILTFNNKKYSHKFKDKVCTPSHPGSMPTVLARVLGQAPPEFG 407

Search completed: October 12, 2001, 16:05:44
 Job time: 187 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

UM protein - protein search, using sw model

Run on: October 12, 2001, 16:05:18 ; Search time 42.34 seconds

(without alignments)
1709.278 Million cell updates/sec

Title: US-09-729-454-2

RefSeq score: 2908

Sequence: 1 MSSNTMLQKTLTLISFSV.....IHPPDHVIGNINFLNYIC 547

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum hit seq length: 0

Maximum hit seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SPTREMBL_16:
1: sp.archaea:
2: sp.bacteria:
3: sp.fungi:
4: sp.human:
5: sp.invertebrate:
6: sp.mammal:
7: sp.mmc:
8: sp.organelle:
9: sp.phage:
10: sp.plant:
11: sp.rodent:
12: sp.unclassified:
13: sp.vertebrate:
14: sp.virus:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	314	10.6	160	4	Q9NXP5	Q9npx5 homo sapien
2	119.5	4.1	860	14	Q9YW26	Q9yw26 melanoplus
3	117	4.0	744	11	Q9FNS6	Q9fns6 mus musculu
4	117	4.0	792	4	O60272	O60272 homo sapien
5	107.5	3.7	1072	10	Q9MOM8	Q9mom8 arbidopsi
6	104.5	3.6	1253	10	Q9SCZ3	Q9scz3 arbidopsi
7	104	3.5	801	10	O04408	O04408 plism sacti
8	100.5	4.5	765	14	Q9JGS8	Q9jgs8 tt virus
9	100.5	3.5	770	14	Q9DHA8	Q9dha8 tt virus
10	100	3.4	527	2	Q9FK09	Q9fk09 chlamydia m
11	99	3.4	654	3	Q9RLP9	Q9rlp9 mycoplasma
12	98	3.4	749	3	O12043	O12043 saccharomyc
13	98	3.4	761	1	O59046	O59046 methanococ
14	98	3.4	817	3	P78875	P78875 schizosacch
15	97.5	3.4	594	10	Q9M8K3	Q9m8k3 arbidopsi
16	97.5	3.4	1049	3	O13824	O13824 schizosacch
17	96	3.3	1063	3	O13983	O13983 schizosacch
18	96	3.3	1103	10	Q9SUK3	Q9suk3 arbidopsi
19	95.5	3.3	765	14	Q9JGS5	Q9jgs5 tt virus

20	95.5	3.3	823	5	Q23141	Q23141 caenorhabdi
21	95.5	3.3	1352	10	Q9ZUF0	Q9zuf0 arbidopsi
22	95	3.3	328	5	Q97317	Q97317 plasmodium
23	95	3.3	1148	5	O17517	O17517 caenorhabdi
24	95	3.3	1714	5	O96326	O96326 plasmodium
25	94.5	3.2	368	2	O47903	O47903 fusobacteri
26	94.5	3.2	677	4	Q9NFB8	Q9nfb8 homo sapien
27	94.5	3.2	999	5	Q9U5A3	Q9u5a3 plasmodium
28	94.5	3.2	4981	5	O77372	O77372 plasmodium
29	94.5	3.2	450	2	Q9EZ92	Q9ez92 lactococcus
30	94	3.2	531	5	Q9UBR4	Q9ubr4 drosophila
31	94	3.2	1440	5	Q9V3S5	Q9v3s5 drosophila
32	94	3.2	1598	4	Q9NM60	Q9nm60 homo sapien
33	94	3.2	1914	4	O13129	O13129 homo sapien
34	93.5	3.2	227	2	Q9PW45	Q9pw45 campylobact
35	93.5	3.2	615	11	Q9JMB9	Q9jmb9 mus musculi
36	93.5	3.2	755	14	P89107	P89107 saquaro cac
37	93.5	3.2	823	10	Q9FGW3	Q9fgw3 arbidopsi
38	93.5	3.2	2554	2	O30981	O30981 bacillus su
39	92.5	3.2	477	10	Q9LWY0	Q9lwy0 oryza sativ
40	92.5	3.2	502	2	Q9KRE5	Q9kre5 bacillus ha
41	92.5	3.2	686	11	O60845	O60845 mus musculi
42	92	3.2	1269	8	Q9TM34	Q9tm34 cyanidum c
43	92	3.2	2643	5	O01552	O01552 caenorhabdi
44	91.5	3.1	647	14	O90140	O90140 buzura supp
45	91.5	3.1	2051	5	O44328	O44328 hiruudo med

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	160 AA.
Q9NXP5	Q9NXP5			
AC	Q9NXP5			
DT	01-OCT-2000 (TRENBLrel. 15, Created)			
DT	01-OCT-2000 (TRENBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TRENBLrel. 15, Last annotation update)			
DE	CDNA FLJ20127 FIS. CLONE COL06176.			
OC	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLON;			
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,			
RA	Okita T., Tanaka T., Nakamura Y., Isogai T., Ohyashi M., Nishi T., Shibahara T.,			
RT	"NEDO human cDNA sequencing project."			
RL	Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AK00134; BAA9065.1; "			
SQ	SEQUENCE 160 AA; 18466 MW; 113114CA43912208 CRC64;			

Query Match	10.84;	Score 314;	DB 4;	Length 160;
Best Local Similarity	41.44;	Pred. No. 1e-18;		
Matches 56;	Conservative 28;	Mismatches 50;	Indels 4;	Gaps 2;
QY	232	IDVTNCKRE-KIEFTQVGNKPPVGGYTLGKWIITTFPNQVOLDTIKINGCLKGLIY 350		
DB	7	ISAKNCKETVAMKECKCFGWTSTIPSGVHMTNPNVCS--LATVMEKCELRGLIY 63		
QY	351	ILGISTLRQMIYFFPKVYTKLPEDHERGIFKKHLLDAERHTOIOMKXSYFFPFFOL 410		
DB	64	LMGISTLRQMIYFFPKVYTKLPEDHERGIFKKHLLDAERHTOIOMKXSYFFPFFOL 123		
QY	411	YSLIDHDYIPREIDRLSGDK 430		
DB	124	YSVKEWEYLRALDRDTGSEK 143		
RESULT	2			

09YW26
ID 09YW26 PRELIMINARY: PRT: 860 AA.
AC 09YW26;
DT 01-MAY-1999 (Tremblrel, 10, Created)
DT 01-MAY-1999 (Tremblrel, 10, Last sequence update)
DT 01-MAR-2001 (Tremblrel, 16, Last annotation update)
DE ORF MSV067 PUTATIVE mRNA CAPPING ENZYME LARGE SUBUNIT HOMOLOG
DE (VACCINIA DIR), SIMILAR TO SWP:P20979.
GN MSV067.
OS Melanoplus sanguinipes entomopoxvirus (MSPV).
OC Viruses; dsDNA viruses, no RNA stage: Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=61191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TUCSON.
RX MEDLINE=99102612; PubMed=9847359;
RA Alonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus.";
RL J. Virol. 73:553-552(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TUCSON.
RA Alonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF063866; AAC97815.1;
SO SEQUENCE 860 AA; 101503 MW; 30420ACBE1037692 CRC64;

Query Match 4.1%; Score 119.5; DB 14; Length 860;
Best Local Similarity 18.3%; Pred. No. 0.25;
Matches 132; Conservative 96; Mismatches 207; Indels 285; Gaps 31;

QY 4 NTML---OKTLILISFVVYTMPIIISQNT---KIMSAL-----NLSISVHYMNSAK 52
DB 126 NMNIIIPKKILMSVNTIHLISIIYPIINGNVYEFLLKTLADNNKMLQYTLN--- 182
QY 53 SLFPTSLIPKPLETELRIKEIEKIDQIIPRPFT-----VN-----TTTSA 98
DB 183 -----VLTQKQFTSYIEI-EIINKLSKDEFFYRQITAFRYIYSINNINPALSTKNI 234
QY 99 THSTATILNPROTYCGDQLDLEVRDHGQKQYGDGFLAAMSSPALTAGASQKWD 158
DB 235 SPSTVTFMPLPDSL-----YLMDSKY-----FLTSKIDGEEVQ 268
QY 159 FNNGTIVLSFTLPMGQVLSLLIHPSEGASALMARQGYDKITPKGFVNGTSH--- 215
DB 269 FTVKNIGIDILV-----VFTECGLT-----LNSNAELCEYLDRODEAFYCKMPQHMP 252
QY 216 -----VFTECGLT-----LNSNAELCEYLDRODEAFYCKMPQHMP 252
DB 299 KQGEYKIDNVKTYIPFVTEIYSTDSKQKTYREMIKFNQINIKNEFDTFKHKPI 358
QY 253 -----FALTYMTRRREVS---YLDKENSLEHRSKY----- 281
DB 359 IPIEKDIYNEALAYKSDKEKITGVLDDKESNIDYLLKIDNTVDYIASLTFHKSPIKI 418
QY 282 ---GVEM-----WKDKHIDVTNCKREKIEETCGVKKRPVQGYTLGGKITTEFCNOV 333
DB 419 HKDGIYTFNLYNDEKN-----ITELLAVEKE-----SNDLITYONDI 457
QY 334 QLDTIKINGCLGKLIYLLGDSILRQVIYFPKYV-----KITLFPDLNETHIEFKHLL 368
DB 458 NMLKFKNIHNRGKNILYAPLCGIEVSFLKSKYINIMLKITTKFRKSYIYNSLK-VIL 516
QY 389 DAERHTQIOMKKHSYFVTPOLYSLIDHDIYIPREDRLSGDKNTAIV-----IT 437
DB 517 DS-----KTFHEKIVPDLTVG-IDVLYTFKINNEYIDKKNLLINONVEKYKPEIE 567
QY 438 FGHFRFPIDIFIRRAIGVQKAIERLFRSPATVVIKENTREHNTETFGDPFGYI 497
DB 568 NMOOORP-PLNIF-----TNLIKTEAISAAKSLCEIIPNHNVLSDIGRGDINKYI 619

QY 498 HYLIH-----KDI-----FKCINVSILH----- 679
DB 620 YICISMLGTDPTDISALSEAGEEYNSLOTRKRVNSTYFKSSLNISLDHLYLKRVSY 679
QY 516 -----ANDMTI-ANGDTIHPPPHVI-----GNCI-NMFLN 544
DB 680 MTOQKIKIFGLIEQWMAIHYSYVDTKKILNLIKLSIDGTGTVIITCLNCKIKNLLN 739
RESULT 3
ID 09ESN6 PRELIMINARY: PRT: 744 AA.
AC 09ESN6;
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel, 16, Last annotation update)
DE NURAL ACTIVITY-RELATED RING FINGER PROTEIN.
GN NARF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR MICE;
RA Ohkawa N.;
RT "cDNA sequence encoding Neural Activity-Related Ring Finger Protein
RT (NARF).";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB043550; BAB17634.1;
SO SEQUENCE 744 AA; 81445 MW; 80336DAZEPEFEFC CR744;

Query Match 4.0%; Score 117; DB 11; Length 744;
Best Local Similarity 28.5%; Pred. No. 0.33;
Matches 41; Conservative 23; Mismatches 50; Indels 90; Gaps 325
QY 67 TETELRI-KEIIEKL---DQQIPRP-----FTVNTTSA 98
DB 266 TETELVLRKQSEKLNELADDPPLHPRENQDLFIETEGIKKSYHNIGITTNVA 325
QY 99 THSTATILNPROTYCGDQLDLEVRDHGQKQYGDGFLAAMSSPALTAGASQKWD 158
DB 326 SEYVATGEGIKQITII-GQPMSTVITTKDKGELCTGNVYLTALSTPQASV-ALGSEILD 364
QY 159 FNNGTIVLSFTLPMGQVLSLLIHPSEGASALMARQGYDKITPKGFVNGTSH--- 215
DB 384 NKNQTYEFLIYQKESDFTLSRL 407
RESULT 4
ID 060272 PRELIMINARY: PRT: 792 AA.
AC 060272;
DT 01-AUG-1998 (Tremblrel, 07, Created)
DT 01-AUG-1998 (Tremblrel, 07, Last sequence update)
DT 01-MAR-2001 (Tremblrel, 16, Last annotation update)
DE KIA0517 PROTEIN (FRAGMENT).
GN KIA0517.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).


```

QY 42 ISVHWNNSAKSLFPTSLIPLKP--TETELRIKEILEKLDQIIPRPPTHVNTTSATH 100
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 LEAHFGKILREBELGKSSI--TRPILRLNVLFRKRLVVLVDVCKPL----- 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 101 STATILNRPDTYGRSD-----QDILILEYRD-----HLGQRKQYGDGF 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 DAESFLGDFMFCPSGLIITTSRKQVFSICRVDOIVYEPPLNEEALQIFSCAFGEKI 334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 139 LRAKSSPALTAGAGKVDNPNNGTYVSPFLTEWGOVSLS---LLLIHPSBG---ASA 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 335 IHESLQK-----LSKKVIDYANGNPL--ALIFGCHSRKNPKPIEIAFPVKYKLAHE 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 LMRARNGYDKI-----IFKGFYNGTSHVFTGCG---LTLNSNAELC---231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 386 IHDAVKSTIDSLSSNEKNFIPLDIACLFGEVNDVCVHLLBEGCGFFPVEINVLVEKCLVS 445
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 232 -----EYLDNR-----DOEAFYC-----244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 446 MAEGRVVNHNLQSIGRKIKNGKRSRLMKPLIITKFLIEDROYLGSDEIAIFIDPSAL 505
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 245 ---MKP-----OHMPCEALTY- 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 506 SFVDNPAFENMKNLRYLKICSSPNGNHVALHPGVKSLPEELRLLHWEHFPPLSLPQD 565
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 258 MTRNREVSYLTDKENSLEFHSKVGVEYEMKDRKHIDVTNCKREKIEETCOYGMKPPYCG 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 FTRNLVILNMC---YSKQLRMEGTKELGMLKRLMLCHSGLQVIOE--LQIALMEY-- 619
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 318 GYLOGKWITTCNOVO-----LDTIKINGCLKGLIYLDSDTLROWIYEPKY 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 620 -IDLOS-----CARLQRFATGHPQLHVLNLSGCIK-----IKSPREV 657
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 368 VKTIKFFPDHETGIFKRLHLLDAERHTQIOWKKSYPFTFO-----LYSLIDHDYTR 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 PPNIEELYXKQTI-----RSIPVTFSPQDMSFYLQHKHKLNR 698
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 422 EIDRLSGDKNTAVITFGQHRPPPID--IFIRALIGVQAIKRLRLSPATK----- 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 699 EV---SSDSOSLSIMVYIDNKLKVLDSGLEIEDIGIAPNKLKYLGGTAIKELPSLMH 755
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 473 ---VIKTEINREMH 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 756 LSELVYLDLENCRLH 771
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7

ID 004408 PRELIMINARY; PRT: 801 AA.

AC 004408;

DT 01-JUL-1997 (TREMBLrel, 04, Created)

DT 01-JUL-1997 (TREMBLrel, 04, Last sequence update)

DT 01-MAY-2000 (TREMBLrel, 13, Last annotation update)

DE ENT-KAURENE SYNTHASE A.

GN LS.

OS *Plasmodium sativum* (Garden pea).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;

OC Fabales; Fabaceae; Papilionoideae; Plisum.

OX NCBI_TaxID=3888;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. TORSDAG;

RA MEDLINE=97260958; PubMed=9107034;

RA Alt-All T., Swain S.M., Reid J.B., Sun T.P., Kamiya Y.,

RT "The *ls* locus of pea encodes the gibberellin biosynthesis enzyme ent-

RT kaurene synthase A".

RL Plant J. 11:443-454(1997).

DR EMBL: U63652; AAB58822.1; -

DR Mendei: 16494; PIsa:1533;16494.

DR InterPro: IPR001906; -

DR Pfam: PF01397; Terpene_synth; 1.

SO SEQUENCE 801 AA; 92717 MW; 5CB88ADE00366844 CMC64;

```

Query Match 3.5%; Score 103; DB 10; Length 801;
Best Local Similarity 19.8%; Pred. No. 5.5;
Matches 124; Conservative 91; Mismatches 107; Indels 214; Gaps 0;

QY 2 SSNTMLQKTLILISFVNTW-----MIFISNFTKMSALNLSIVHWNNSA 51
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 SAHRIINTLACVILR--SWNHSEKCDKQIF--FRNLKLELNNEEHM ICF----- 267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 52 KSLFPTSLIP---LKP---TETELRIKEILEKLDQIIP--PRPPTHVNTTSATHST 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 EVAFP--SLLEFARGIKPLKCPNDSPI-LKNIFERDKELIRIKELIMKVP7----- 255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 103 ATINPMDIYCGDGLDIL--LEPHDHLGQRKQYGDDELRAKSSPALTAGAGKVDNPN 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 -TLHSLEGMSGLDKQLKQSD-----GSL---FSPSSTAFALMLQTKDNG 301
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 GTIYVSPFLTEWGOVSLSLLIHPSGASALW---RAVNOGYDKIIFKGF---VNGISH 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 LKYLNVVKKFNGVVP---NVYPVLFEHIVVDRLERLGISR--FFHREIKLCNVYSK 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 VTEEGCL--TLNSNAELCEYLDRODA-----PY MKP 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 IMSEKICMARSNVQ-----DIDTAAAFRLRLGHQVSAHVFKNFERNDPFCFAG 410
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 248 QHMPG--EALTYTTRNREVSYLTDKENSLEFHSKVGVEYEMKDRKH----- 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 Q---CTQAVTGMYNLFASQVLPFGKILFHAHKSAYLKEKHEANLIDKMLIMKLP 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 292 -----IDVTNCKNREKIEETC---OYGMKPPYCGGTTLOGKAITTEGQVQVLTITNG 342
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 468 EHYGVALDMPWANDRIETRFIDYQASQVWIGTGLFRMAYVNNNNVYIELKLLYNN 527
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 CLKGLIYLDSDTLROWIYEPKYVKTLEFED--LHETGIFKRLHLLDAERHTQIOWK 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 528 CQAOHLI-----EW-----NVQIYTESRLGFGLSKRLLL----- 560
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 SYPFVFOYSLIDHDYIPREIDRLSGDKNTAVITFGQHRPPPIDIFIRKALGVKAI 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 561 AYFLATGSIPE-----PERSHRLAMAKTALLET----- 591
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 462 ERLFLSPATKVIITKTEINREMHIEERFGPHGYHYLLMKITFRQ-----LN 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 592 -----KGYRNEDLRKDF--AKFNDHIDVDRYSIARMKNKTEHVESLFAF 639
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 511 VGLIDAMDTIAYGDTIHPPDHVIG 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 640 IGFI-SMDVRLSYG-----HEIG 656
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8

ID 09JG58 PRELIMINARY; PRT: 765 AA.

AC 09JG58;

DT 01-OCT-2000 (TREMBLrel, 15, Created)

DT 01-OCT-2000 (TREMBLrel, 15, Last sequence update)

DT 01-OCT-2000 (TREMBLrel, 15, Last annotation update)

DE PORFL.

OS IT virus.

OC viruses; ssDNA viruses; unclassified ssDNA viruses.

OX NCBI_TaxID=66887;

RM [1]

RP SEQUENCE FROM N.A.

RA Tanaka Y., Mizokami M., Orto E., Ohno T., Nakano T., Hayashi K.,

RT Identification of a 23kDa protein encoded by putative open reading

RT frame 2 of IT virus (ITV) genotype I different from the other

RT genotypes".

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB030487; BAA90406.1; -

SO SEQUENCE 765 AA; 90225 MW; E845A26B69D9D707 GCF64;

Query Match 3.5%; Score 100.5; DB 14; Length 765;

Best Local Similarity 20.2%; Pred. No. 8.3;
Matches 94; Conservative 74; Mismatches 190; Indels 107; Gaps 23;

```

QY 14 LLSFSVVTWM-----FIISQNTKLMSALNLSIVHYMNSAKSLFPR-TSLPLKPLTE 68
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 264 VTFQVLRSMYNDALSLIPSNFEQ-----DENAOKLYNEISSYPYANTE 309
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 TELRIKELIKLQD-QIPRP--THVNTTSATHTATILNP-----RDYCHGDDLDL 121
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 310 TTAQIKKREVENTEKISTPNPQOSNYVNTITFTTAOSITTTPTTSDMYGYATYKKA 369
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 LHVDRHLGOKKQYGDPLFARRSPALTAGASKVDENNNGTYVSTLFWESQVSLSL 181
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 370 I-TKVLPLAAKLY--ETQTKNLSTFTFGS--EYLEYHGGLY----- 407
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 182 LHPSEGASALMRARNOGYDKITFKCKEYNGTSHVTECGTLNSNALECYL---DDRD 238
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 408 -----SSIMLSAGRSYFET--KGAVTDICYNPTDGG--EGNMLMIDMLSKDSRY 454
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 239 OEAFCCKPQHPMCEALTY-----MTNRREVSYLTUKENSLFHRSKYGVEMMKDKKH 291
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 455 DKARSKCLIEKLPMMAAVGYAECATGDSNI-----DMNARVVMKQPYTVQMD--- 507
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 292 IDVTNCKNREKIEETCOVMKPPVPG-----YTLOGKWITTFPCNOVQDITKINGCLKG 346
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 508 ---TSDPLRGPIYSFNG-KGKMGGTNGOPIRMAKYR--CLFPHQKEVLEAIQ-QSG 560
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 347 KLIY-----LLGDSTLQKQIY-----YFRKYVYKTKFQDLHEGIRKKHLLDAERHTQ 395
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 561 PFAVHSQKKAVALGLKFRFHMIMGPNPFQVYVNRNCKDQGSTPRK-----PRSVQ 613
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 396 IQMKKHSYPFVTFOLYSLIDHYIPREIDRLSGDKMTAIVTFEQ 440
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 614 IIDPKYNTPELTIHAWDFRGFGPKAIKMQOQPTDAELLPRG 658
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9

Q99HAB PRELIMINARY; PRT; 770 AA.

AC Q99HAB; (TEMBLrel. 16, Created)

DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)

DE 01-MAR-2001 (TEMBLrel. 16, Last annotation update)

DE VPL.

OS ORF1.

OT TT virus.

OX Viruses; ssDNA viruses; unclassified ssDNA viruses.

OX NCBI_TaxID=68867;

RN [1]

RF SEQUENCE FROM N.A.

RC STRAIN-VT416;

RX PubMed-11024126;

RA Kamehara T., Hino S., Miyata H.;

RT "Three Spliced mRNAs of TT Virus Transcribed from a Plasmid Containing

RT the Entire Genome in COS1 Cells.";

RL J. Virol. 74:9980-9986(2000).

DR EMBL: AB041821; BAB15948.1; -

DR EMBL: AB041007; BAB15944.1; -

SO SEQUENCE 770 AA; 90165 MW; 1BD91574BED7D602 CRC64;

Query Match 3.5%; Score 100.5; DB 14; Length 770;
Best Local Similarity 19.4%; Pred. No. 8.4;
Matches 94; Conservative 66; Mismatches 171; Indels 153; Gaps 21;

```

QY 30 NFKRLMSAL--NLSIVHYMNSAKSLPPTSLIPKPLTELEIKELIKLQDQIPPR 87
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 265 NFOVLOSMTYQNTSILPDKSEKELLQIWDIYIPYNTTQITLAKLFID-AGNTPTMT 323
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 86 PFT-----HVNT--TISA-----THSTATILNPROTYRG---DQDILILEYRQH 127
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 424 PATWMSYINTTKFTTAATTYTPGTITTYVMTLNCDSWRCGTIVYDKIKLN----- 377
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 128 LGOKROYGDFLARMSSPALTAGASKVDENNNGTYVSTLFWESQVSLSLIHSE 187
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 378 --PQOATLYSKATKTLTGLNFTTDDHTEYHGLY----- 411
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 GASALMRARNOGYDKITFKCKEYNGTSHVTECGTLNSNALECYLDR---DOEAFY 243
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 412 --SSIMLSAGRSYFET--TGAVTDICYNPTDGG--EGNMLMIDMLSKNNYDKVQSK 464
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 CKRQHPMCEALTY-----MTNRREVSYLTUKENSLFHRSKYGVEMMK 287
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 465 CU-ISDPLMAAIVGYEPCAKSGDONIHMAKLLINSPITDQVLY----- 511
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 DKRHIDYTNCKNREKIEETCOVMKPP-----PYPGSYT-----LOGKWITTFPCNOVQ 334
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 512 ---HTDPTK-----GVPYSLNFGNGMKPGSSNVPIMRAKWPPLTFHQE 555
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 335 L-----DTIKNGCLKGLIYLLGDSTLQWMIYFPPKYVTKLAFPLDHTG 380
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 556 VLEALQSGFPAYHSDIKKVSIGMKYFKKIMOGNPNVQVYVRP-----CKETHSG 608
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 381 IFKHLLDAERHTQIQMKKHSYPFVTFOLYSLIDHYIPREIDRLSGDKMTAIVTFEQ 440
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 609 -----NVPRLQIVDEKYNSELTFTWDFRGLGPRAIQPMQOQPTTIDIFSAG- 660
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 441 HFRP 444
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 661 HKRP 664
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10

Q9PKU9 PRELIMINARY; PRT; 527 AA.

AC Q9PKU9;

DT 01-OCT-2000 (TEMBLrel. 15, Created)

DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)

DE 01-MAR-2001 (TEMBLrel. 16, Last annotation update)

DE 4-ALPHA-GLUCANOTRANSFERASE.

GN TC0362.

OS Chlamydia muridarum.

OX Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=83560;

RN [1]

RF SEQUENCE FROM N.A.

RC STRAIN-MOPN / NIGG.

RX MEDLINE-20130255; PubMed-10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heldelberg J.F.,

RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,

RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,

RA Eisen J., Fraser C.M.;

RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia

RT pneumoniae AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).

DR EMBL: AE002303; AAF39223.1; -

DR TIGR: TC0362; -

DR InterPro: IPR003385; -

DR Pfam: PF02446; 4A-glucanotrans; 1.

SO SEQUENCE 527 AA; 61328 MW; 1012A95DA309E2A CRC64;

Query Match 3.4%; Score 100; DB 2; Length 527;
Best Local Similarity 20.0%; Pred. No. 5.3;
Matches 92; Conservative 68; Mismatches 158; Indels 142; Gaps 25;

```

QY 65 PLTELEIRI--KEIIEKLQOI--PPR-----PFTVNTTTSATHTATILNPROTYCR- 114
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 PLTSRLKILNPSIRKYNVNDTSPKIGICVPLFSIHQNSCGIGERLDLIMIDWCL 61
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 -GDQDILILEYRQHLGOKROYGDFLARMSSPALTAGASKVDENNNGTYVSTLFW 173
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 CGQIQLQILPLINDGSCSSPYNS-----ISSISL-----NPJHLSISALPYK 103
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 GQVSLSLLIHPSEGASALMRARNOGYDKII-FKGRFVNGTSHVTEC-GLTINSNAELC 231
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      104 EEVSSSRKLIIDEMORLSQSOV-----NZEKVIIPKRAFE-----KEYFVCKSKMLTNHPDFC 157
Oy      232 EYLDROD-----EAYGCKPOHMCBALYMTNRREVSUTLDEKNSL-----FIRS 279
Db      158 DFC-BREYUMLHNYALFCSIRHNLNLPINHMSTYTTLDSISIHETPAKIDETSLQ 216
Oy      280 KYGVEMMD-RKNIIDVTNCKREKI-----EEGCOV-----GCKPVCQGYTLQ 322
Db      217 YLCEFOCKOYKRNADKSGLLIGDPIILSKDSQDWMEYRKYFSSSESQSP-PDYNAE 275
Oy      323 GK-WITTEPCNOVOYDITKINGSLGKLIYLDGSLROMIYFPKVVKTLL-----FEDL 376
Db      276 GONMYLPIYNN-----KTLRODAYHMMK--ERLPAENFELY 311
Oy      377 ---HETGFGKHLILDAERHTQIQMKKSYFVTOYLSLDH-----DYPRED 424
Db      312 RLIDHYVGLFR-----FWVDELGRGRREPODDPYDIDGSD 347
Oy      425 RLSGDKNALYITFGCHFRPFDIFRRLIGYQAKIERL 464
Db      348 ILSHLKASSMLPIGDLGTTAVD-----VKQLESL 379

```

RESULT	11	
09RLP9		
ID	09RLP9	PRELIMINARY; PRT; 654 AA.
AC	09RLP9	
DT	01-MAY-2000 (TEMBLrel. 13, Created)	
DT	01-MAY-2000 (TEMBLrel. 13, Last annotation update)	
DT	01-MAY-2000 (TEMBLrel. 13, Last annotation update)	
DE	P75 PROTEIN PRECURSOR.	
GN	P75.	
OS	Mycoplasma hominis.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;	
OC	Mycoplasmataceae; Mycoplasma.	
OX	NCBI_TaxID=2098;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=183;	
RA	Mygling T., Birkeland S., Christansen G.;	
RT	"Cloning, DNA sequencing and restriction enzyme analysis of the p75	
RT	gene indicates the presence of a different type of surface protein	
RL	variability in Mycoplasma hominis ";	
DR	submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: AJ243903; CAB6240.1; -	
KW	Signal.	
FT	SIGNAL.	
SO	SEQUENCE	1 27 POTENTIAL. 654 AA; 75607 MW; 38E/C46FF5D44091 CRC64;

Query Match 3.4%; Score 99; DB 2; Length 654;
 Best Local Similarity 18.5%; Pred. No. 8.9;
 Matches 103; Conservative 98; Mismatches 188; Indels 168; Gaps 28

```

OY      2  SENTLOKTLILIPSVYVMMI--FIISQFRIKMSALNISVHYVM--NSAKSLPEKT 56
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      122 SONDLKK---LAVELIETITKIMMLISSNEK-----NISEKEANILINAKNALEKS 173
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY      59 S-----LIPKPLTETELRAKELIEKDOQIPRPPTHVNTTSATSHSTATIINPRDYCR 114
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      174 NQOEYKQIKRQVNETINKLEEISINLEKGRKSEPTTKLE-----LREKLTWYL 222
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY      115 GDQL-----DILLEVRHDLGQRORQYGDGFLRARRSSPALTAGAGKVDPMFNNGYIVLSEF 168
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      223 DSENTSEFAYMDLAEIKELIQOENV-----SIVSNEKEIKSAISOVD----- 266
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY      169 TLEMEGVQSYSL--LIH-----PSEGASALMARNOGYDKLIIFKQFVN-----OTS 214
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      267 -LEWIANIKKSIIDKQFLYLENKKPSLDVKFLAHLDELSQLATVIFENIETTERBOL 325
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY      215 HNFTEGGLTNSNAELCEIYLDROEAFYQKRPQHPCEALTYWTTRNREVSYLTDKENS 274
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

```

Db 326 NIKTILTRISANVEKONIEE-----VNINS 46
QY 275 LPIRSKVEEMKRRKHT---DYTNCKRKEIEICQVMKRPVGSYTLCKRKHITP 41
Db 355 LUKKYNINNKNKOQIEISGDIJOLUKKQIE-----NNVHIVKVLKLN 42
QY 332 QVOLTPIKINGCKLKIYL--LGDSTLROMIYYPKVKIKLFFOLHEIGIRKHLID 40
Db 403 KA--NTIKEGSFAKKAILLETLRD-TISELFYDLTHLRLEFFIKNNIDIK--LLP 47
QY 390 AEHHQIQOMKHSPPVLTQSLIDHDYIPRIOLSLCKNT---AVITGCGHPRITP 44
Db 457 ASDQRTI-----NNYL-NKIKIT-ENNSQTPQSEITFEYSAF- 44
QY 447 IDIFIRAIQVOKAIRLP-----LRSPATVILK-TENIRMEHIEFERGCH- 44
Db 495 -----KQSEIFRIKSSDLASADVNIKELTNLKE--IKKKNYDNLVOT 54
QY 495 GYTHIKIDIFKDYLV 511
Db 541 SYLKEKIEETIKNEST 557

```

RESULT	12			
ID	012043	PRELIMINARY:	PRT:	749 AA.
AC	012043:			
DT	01-NOV-1996 (TREMBLrel. 01. Created)			
DT	01-NOV-1996 (TREMBLrel. 01. Last sequence update)			
DE	01-MAR-2001 (TREMBLrel. 16. Last annotation update)			
GN	CHROMOSOME XV READING FRAME ORF YOR0818C (YOR29-12 PROTEIN).			
OS	YOR2964C OR YOR29-32 OR YOR0818C.			
OC	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID:4932:			
RN	[1]			
RP	SEQUENCE OF 434-749 FROM N.A.			
RA	Bohn C., Bolotin-Fukuhata M., Dalgan-Fornier B., Dang D.V.,			
RA	Valens M.,			
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Voss H., Benes V., Rechmann S., Teodoru C., Schwager C., Paces V.,			
RA	Ansozge W.;			
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	MIPI.			
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RA	Benes V., Andrade M.A., Rechmann S., Teodoru C., Baureyl A.,			
RA	Sander C., Valencia A., Ansozge W., Voss H.;			
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	SEQUENCE OF 435-749 FROM N.A.			
RA	MEDLINE:97279235; PubMed:9133743;			
RA	Valens M., Bohn C., Dalgan-Fornier B., Dang V., Bolotin-Fukuhata M.;			
RL	"The sequence of a 54.7 kb fragment of yeast chromosome XV reveals the			
RL	presence of two tRNAs and 24 new open reading frames.";			
RT	Yeast 13:379-390(1997).			
DR	EMBL: 274989; CAA99274.1;			
DR	EMBL: X94335; CAA61004.1;			
DR	EMBL: 270678; CAA94506.1;			
DR	SSD: S0005607; YOR0818C.			
SO	SEQUENCE 749 AA: 84715 MW; F3C5DBA97B6BA5B8 CDDC44;			

Query Match 3.48; Score 98; DB 3; Length 749;
Best Local Similarity 20.48; Pred. No. 13;
Matches 120; Conservative 70; Mismatches 224; Indels 174; Gaps 27;

```

QY 5 TMLQKLLILISFVYVWMIIFISQNFTKLSALNLSIVHWNNKAKSLFPTSLILEK 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 TLRKOLL-----SDLSQKKAHSIYQOMNDIASLIDLITGLSSEK 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 PLTELERIKEILEKDOIIPRFPHVNTTSATSHSTATILNPDYTCNGDQDILLEV 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 TIDESSLYNKLLDOLLIRM-----RLRLRTYD-VHRLLYLIRTKWVRNLGNMNNLYR 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 RDHLGQRKQYGDPLFRANSSPALTAGASKYMDNNGTYLVSTFLWEGQVSLILLIH 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 HSHTG-TKQIINDYLEE--SQAVLTA-----LIH 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 PSE-----GASALMRARNOGYDKIIFPKGVNSTSHVTEGGLTNSNAELCEYLD 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 QSMNNDHYLLGLQOFR-RNIGRTALVLSGGSTFGLFHIGVLAAL--FESDLMKXVLSG 214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 EQ-----EAFYQKPMPOHPCFALTYMT--TRNREVSYLTDKNSLIFRKSXVGEKKKKH 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 SSAGATVASFV--HTQELPSLLTNVLMNEFIEND-DNS-----253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 292 IDVTNCKRREKLEETQVQGMKPPYGGYTLQCKWITTFQNOVOLDTIKINGLCKLYL 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 -KSPENLLIKISFCQNG-----TWNNQPLINT-----MLSF 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 352 LGDSTLRO-----WIYEPKVVYKTLKFFDL-----HETGIFKKH 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 LGMLTFEAYNKTGKLINTVSPASIEQPKLNNLAPVNLVMSAVCASGLQVPEPT 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 386 ILDAERHTQIQMKHSPVTFQLYSL-----IDDDYIPREIDLSDCKMTAVITFG 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 PLFEKDPHT--GKIKKQATNHLNMMFMGSDVND--MPISLSEMFVDDIACQ 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 440 QHFRPPDIDIFIRRAIG--VQKAIERLFKSPATKV--IKTENIEMHIEITERGQPHG 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 401 VNIHVFPLKFSNTQVGEIEKEITARF-RNQQVTKIKFSPDETIFHDILKEL--EFH 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 496 YIHVLMKDIFKDLNGLIDAMDMITAVGDTIHPRHVGNQINNEL 543
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 YL MTKLKHFLQOYSG-----NTILPDLMSVQGFHEVLKNSQLFL 499
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13

```

ID Q59046 PRELIMINARY: PRT: 761 AA.
AC Q59046;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN M01652.
GN M01652.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae:
OC Methanococcus.
OX NCBI_TaxID=2190;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 26611.
RX MEDLINE=96337999; PubMed=8688087;
RA Ruit C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
KA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
FA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
KA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
KA Uiterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
KA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
KA Klepek H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RJ jannaschii."
RL Science 273:1058-1073(1996).
CC 1. SIMILARITY: TO PROKARYOTIC TYPE I/III TOPOISOMERASE FAMILY.
DR EMBL: U67605; AAB9673.1;
DR HSSP: P06612; IECL.
DR TIGR: M01652;

```

```

DR InterPro: IPR00380;
DR InterPro: IPR002936;
DR Pfam: PF01131; Topoisom. bac. 1.
DR Pfam: PF01396; zf-C4_Topoisom. 3.
DR Pfam: PF01751; Toprim. 1.
DR PRINTS: PRO0417; PRTPRSMRASET.
DR PROSITE: PS00396; TOPOISOMERASE_I_PROK; UNKNOWN_1.
DR SMART: SM00493; TOPRIM. 1.
KW Hypothetical protein: DNA-binding; zinc-finger.
FT ZN_FING 600 648 C4-TYPE (POTENTIAL).
FT ZN_FING 680 706 C4-TYPE (POTENTIAL).
FT ZN_FING 711 747 C4-TYPE (POTENTIAL).
FT ACT_SITE 315 315 DNA_CLEAVAGE (BT SIMILARITY).
SQ SEQUENCE 761 AA; 87832 MW; 2F9C95735E20D82 CRC64;

```

Query Match 3.4%; Score 98; DB 1; Length 761;
 Best Local Similarity 21.5%; Pred. No. 13;
 Matches 53; Conservative 40; Mismatches 76; Indels 78; Gaps 13;

```

QY 253 EALTYMTNRNREVSYLTDK--ENSLFHRKXVGEEM-KDKKH-DVTNCKRREY----- 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 EAVSYKRIKPKETQRIQAKLYENALISYRITSQKLPDKRKYLEDILINRHPYQKNA 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 EETQVQGMKPPYGGYTLQCK-----WITFC 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 348 ERIKKNLKP-----VEGKREDPAHAIHVIDIPKELSEKEKEIYDILARRITLAATP 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 331 NOVQDLTIKINGLCKLYLIGDSTLRO-W-IYEPKVVYKTLKFFDLHTGIFKKHLL 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 401 DNRREYLVNKKIDIKGEKFLSSRYVEGHEHLYPK-----FELILPPLKKNKI 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 LDEERHTQIQMKHSPVTFQLYSLIDDDYIPREIDRLS-GDKNTAVITFGQHFRPP 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 IKVEKLTITR--RKETOPPKRYTVASII-----KELEKRLGTAKMTAREI----- 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 447 IDIFIR 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 IDKLIR 502
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 14

```

ID P78875 PRELIMINARY: PRT: 817 AA.
AC P78875;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE TREHALOSE-PHOSPHATASE (EC 3.1.3.12) (TREHALOSE 6-PHOSPHATE
DE PHOSPHATASE) (TPP).
GN TPPI OR SPAC19G12.15C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes:
OC Schizosaccharomycetales; Schizosaccharomycetaceae.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Oliver K., Harris D., Wood V., Bartell B.G., Rajandream M.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=974;
RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 505-817 FROM N.A.
RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioaka S., Kato K., Nakai K., Okayama H., Nojima H.;

```

*Identification of open reading frames in Schizosaccharomyces pombe
RT CDNA3.7;
RL DNA Res. 4:363-369(1997).
CC -1- CATALYTIC ACTIVITY: TREHALOSE 6-PHOSPHATE + H(2)O -> TREHALOSE +
CC ORTHOPHOSPHATE.
CC -1- SIMILARITY: CONTAINS A DOMAIN FOUND IN ALL THE PROTEINS OF THE
CC TPS COMPLEX.
DR EMBL: AJ242743; CAB45142.1; -
DR EMBL: Z97209; CAB10126.1; -
DR EMBL: D89225; BAA13886.1; -
DR InterPro: IPR001830; -
DR InterPro: IPR003337; -
DR Pfam: PF00982; Glyco_transf_20; 1.
DR Pfam: PF03356; Trehalose_Phase; 1.
KM Hydrolase.
FT DOMAIN 1 547 TPS COMPLEX DOMAIN.
SQ SEQUENCE 817 AA; 93877 MW; 23CFD2533505CBA CRC64;

Query Match 3.4%; Score 98; DB 3; Length 817;
Best Local Similarity 20.1%; Pred. No. 15; Mismatches 114; Indels 102; Gaps 15;
Matches 66; Conservative 47;

QY 287 KDRKHIDVT--NCKREKIEETCOVGMKPPVPGVTL-----QCKWITTCNQVO--- 334
DB 113 QDVKRFETYSQWNSGERSTERYVWMLPGEKSEITINETSQGSRLMAEYENIRPLI 172
QY 335 -----LDTIKINGCLGKLL--IYLLGDSLTRO--WLYYPRKVVYTKL 372
DB 173 HYKWPSEYDENEBQWMDYVKNNAHAFADKICEIKYKGFIIYQDYSLFVPLIRNK- 231
QY 373 FFDLHETGIFKKHLLDAE-----KHTQIQMKKSTPEYTFQYLSL----- 413
DB 232 -IDDAVIGFHHHPSPSEIARCPRRALILRSYLGADFIQFDYSAHFSCSYLD 290
QY 414 --IDHDYIPREIDRLSGDKNTALVITFGHFPPIDI--FIRRA--IGVKAIEPLFLR 467
DB 291 LEIGHDMV-----NLNGNKVYRAITVG-----IDVPRLINSGNVSYSKLEELNKR 338
QY 468 SPATKVIKNTENIREMHIETFRGDFHGYI-----HYLINKDI 505
DB 339 YENMKVILGRDLDELGYVQKLSFORLRTPEMKKKVYLIIQITISAFKHKLLSTI 398
QY 506 FKDLNGLIIDAMDNTIAGTDTIHPDHY 534
DB 399 KK-----LVQA--INQEGTDDYTPVHHV 420

RESULT 15
Q9MRK3
AC Q9MRK3: PRELIMINARY; PRT: 594 AA.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE F28L1.9 PROTEIN.
GN F28L1.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Beutlo M., Greasy T.H., Haas B., Wu D.,
RA Rowling C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F28L1 genomic sequence.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC018907; AAF30308.1; -
SQ SEQUENCE 594 AA; 67574 MW; FDA7ECA139C5CC3C CRC64;

Query Match 3.4%; Score 97.5; DB 10; Length 594;
Best Local Similarity 18.7%; Pred. No. 10; Mismatches 128; Conservative 83; Indels 407; Gaps 17;
Matches 128; Conservative 83;

QY 11 LLLIST-SVYWMIFLIS-QNFKLM--SALNSTSVYWNNSAKSLIKSLILKPL 66
DB 31 LTALVCFMLVMSIDGCSIOSFQVQPRFNAVSRLS-----PSSPMSKTNLNSKPH 66
QY 67 TETELRKEIIEKLDQIIPRPPT--HVNTTSAT-----HSTAI-----LN 107
DB 87 R-----QNLTMARNLVPPKNTLITSNSRYVFEVITGSKNFTANLKMOMLA 116
QY 108 PRDTYCR-----GDQDLLEFVRDHLQCKQVGFPAHMS 146
DB 137 PGAPCRKAEKVEISVPGVDISVELTGEIHEFKFQAIDESKNVCIQGYFFTDS- 195
QY 146 PALTAGASGK---VMDFNQTYIVSFTLFEW--GVQSLT--LLIHSEANALBRANQ 198
DB 196 ----GENKKSPPYKDKGNGTYSFSLQVHPFADPRLIYLFRHYQK--FSTSL 248
QY 199 GYKLIIFKGFNGTSHTVTECGLTLSNAELCEYLDROQEAFCMKPQHPCEALTYM 258
DB 249 GFDR-----KLRN-----VALPEV 292
QY 259 TTNKREVSYLTDKENSLEHRSKYQVEMKDRKHIDVTNCKKREKIEETQVGS- 310
DB 263 KTPDYVLPELRSCKSKDFNRDAMSGWTLGKN-----DEQSLSDGRTCL 309
QY 311 -----MKPPVPG--GYTLQKMT--TFPCNOVOLDTIKINGCLGKLYLLGDS- 355
DB 310 AADFPCKRMCQGAVALIESNGVAISSHCSFKLSAKRWDCLKGKIFFMGDSNHYDSI 369
QY 356 -TLRWIYIFPKVYVTKLKFPL-----HETGIFKKHLLDAEHRITQIQMKKS 402
DB 370 RNLNLVGLHPEIIPVPRRDMKFSNPKPSSEYRITISIFNGH-----WMEYK 417
QY 403 YPEVFO--LYSLIDHYIPREIDRLSGDKNT--AIVITRQD-----PRP- 445
DB 418 ---NVQGLDSLKDRDF--RELLKTYFNEETINRPDMYINSGLDGJHMTSLKATFKGA 471
QY 446 -----PIDIFIRAIQVOKAIEKLEFLASPAKVYI1KTENIPEMHJ 486
DB 472 ETAAAFKREVEVDGKSGLOPPEVIFKNTIAT--GGYAFMLAFNPS-----KMEAFNVFL 525
QY 486 ETFRGDFHGYIHYLINKDIFFDLNGLIIDAMDNTIYVG-----TDITR----- 541
DB 526 EK-----MRD--AGLVTSVQDNFDMITPYMINDRNCNGVHYGRALAKMKKH 599
QY 532 DHVIGNQ-----INMELNYIC 547
DB 570 DGEIGHQYFVDMLVHVLNALC 592

Search completed: October 12, 2001, 16:05:22
Job time: 165 sec

Fri Oct 12 16:09:19 2001

us-09-729-454-2.rspt

Page 9

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2001, 16:02:37 ; Search time 29.42 Seconds
(without alignments) 1229.875 Million cell updates/sec

Title: US-09-729-454-1

Perfect score: 2485
Sequence: 1 MKISMINYKSLALLFIHAS.....VHPQHVYQNOINILNYIC 475

Scoring table: BLCSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: 1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1606	64.6	540	2 B45665	adult-specific 61.
2	127.5	5.1	792	2 T00082	hypothetical prote
3	119.5	4.8	3603	1 D69681	peptide synthetase
4	114.5	4.6	807	2 A86740	teichoic acid bios
5	109	4.4	688	2 T33709	hypothetical prote
6	104.5	4.2	4563	1 LPHUB	hypothetical prote
7	104	4.2	694	1 VGBPH6	apolipoprotein B-1
8	103	4.1	793	2 S73779	glycoprotein H pre
9	101.5	4.1	1902	1 B44858	leucine-tRNA 11ga
10	101.5	4.1	3649	1 S18268	lactocarpin (EC 3.4
11	101	4.1	988	2 T52134	delta-(L-alpha-am
12	99.5	4.0	1507	2 T42631	zwille protein 1m
13	98.5	4.0	795	2 T07709	breast cancer tumo
14	98.5	4.0	1004	2 A71617	hypothetical prote
15	98	3.9	470	2 S54069	SEPA antigen/papal
16	98	3.9	657	2 D81308	hypothetical prote
17	97.5	3.9	634	2 T08610	probable integral
18	97	3.9	538	2 S52472	probable ABC-type
19	97	3.9	905	2 S53429	cell fusion protei
20	97	3.9	1366	2 S57664	hypothetical prote
21	96.5	3.9	1802	2 J71616	1ga-specific metal
22	96.5	3.9	774	1 J00550	hypothetical prote
23	96.5	3.9	1245	2 S51255	1,4-alpha-glucan b
24	96	3.9	770	2 S76095	probable membrane
25	96	3.9	2692	2 T23768	hypothetical prote
26	95.5	3.8	1102	2 T28104	hypothetical prote
27	95.5	3.8	1685	2 T02750	probable DNA-direc
28	95	3.8	610	1 J04383	acetyl-CoA carboxy
29	95	3.8	819	1 T83350	3'-phosphadenosin
					outer membrane ush

30	95	3.8	1021	2 T08601	hypothetical prote
31	95	3.8	1328	1 S04273	retrovirus-related
32	94.5	3.8	740	2 T10619	hypothetical prote
33	94.5	3.8	803	2 T00092	hypothetical prote
34	94	3.8	971	2 A70179	exodeoxyribonucle
35	93.5	3.8	650	2 T41681	hypothetical prote
36	93.5	3.8	1209	2 T31657	probable serine/th
37	93.5	3.8	1311	2 T33757	reverse transcript
38	93.5	3.8	1616	2 S62504	hypothetical prote
39	93.5	3.8	1958	2 T39808	conserved hypothet
40	92.5	3.7	2894	2 C64474	hypothetical prote
41	92.5	3.7	368	2 S46435	hypothetical prote
42	92.5	3.7	584	2 H71513	porin precursor -
43	92.5	3.7	1462	2 T50338	probable sdsna exo
44	92.5	3.7	1902	2 B45764	probable guanin-n
45	92	3.7	462	2 T19231	lactocarpin (EC 3.4
					probable cysteine

ALIGNMENTS

RESULT 1
B45665
adult-specific 61.9k brush border protein precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 03-May-1994 *sequence_revision 03-May-1994 *text_change 04-Mar-2000
C:Accession: B45665
R:Bohl, W.; Schmid-Chanda, T.; Semenza, G.; Mantel, N.
J. Biol. Chem. 268, 12901-12911, 1993
A:Title: Messenger RNAs expressed in intestine of adult but not baby rabbits. Isolates
A:Reference number: A45665; MUID:93286138
A:Accession: B45665
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-540 <BOL>
A:Cross-references: GB:212840; NID:q1761; PIDN:CAA78302.1; PID:q1762
C:Superfamily: rabbit adult-specific 61.9k brush border protein
C:Keywords: Intestine

Query Match	64.6%	Score 1606	DB 2	Length 540:
Best Local Similarity	57.3%	Pred. No. 7.4e-115:		
Matches 309; Conservative 67; Mismatches 91; Indels 72; Gaps 2:				
QY	8	YKSLALLFLASWIIIFVFNQSTKYSALNTSLHYNNSTKSLFPTPLISKPLTE 67		
DB	3	HKYIKLICLLAICVLCIISNSTKIKGALKIPNSHYNTSMISSIFKMSVSYKSLTE 62		
QY	68	TELRIKRIIEKLDQIIPRPPTHVNTTTSATHTATILINPDPTYCRGDLITLLEVRDL 127		
DB	63	TELKVEIKLEKLDRIIPRPPTHVNTTTSATHTATILINPDPTYCRGDLITLLEVRDL 122		
QY	128	GRKQKGGFLFARSSPALMAGASQKVDFFNKGIVYLSFTLFWEGOVLSLIIHPSEG 187		
DB	123	GHQKQGGGFLFARSSPALMAGASQKVDFFNKGIVYLSFTLFWEGOVLSLIIHPSEG 182		
QY	168	VSALMSRNQGDRIYIFTGQFVNGTQVSEEGILITNMLCOYIDNRDQEGYCYRPQ 247		
DB	163	ASALMRNRQGDRIYIFTGQFVNGTQVSEEGILITNMLCOYIDNRDQEGYCYRPQ 242		
QY	248	HMPCALITMYKKNKVSYSKQESLFRSNVGEIEMKFNITISKCN----- 297		
DB	243	HMPCALITMYKKNKVSYSKQESLFRSNVGEIEMKFNITISKCN----- 297		
QY	298	TLKSVDLHESGKLOHOLAVDLDRNINIMQKTCYPLIGSMYSYKEMEYLTFRADRTG 356		
DB	362	TLKSFEDLHETGNFKKILLDAEKHQTQIOWKKSHRPVYQLPSVHDIGYIPQETDRLIG 421		

QY 400 MWIKTENREMYNDAREFSDPHGYOYLIIKDIFDLSVSIIDAMDTIATG 452
 DB 1749 -VPITKINKRFRVSGGEJG-----THVSR--FTELLPNV-----SITSYG 1788

RESULT 4

hypotheical acid biosynthesis protein [imported] - Lactococcus lactis subsp. lactis (strain G130)
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
 C:Accession: A86740

K. Bolotin, A. Mincker, P. Manger, S. Jallion, O. Malarme, K. Weissenbach, J. Ehrlich
 Genome Res. in press, 2001

A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625

A:Accession: A86740

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-807 <STO>

A:Cross-references: GB:AF005176; MID:q12723855; PIDN:AAK05019.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: ladh

Query Match 4.6%; Score 114.5; DB 2; Length 807;
 Best Local Similarity 20.0%; Pred. No. 1.1;
 Matches 90; Conservative 76; Mismatches 175; Indels 109; Gaps 18;

QY 29 NSTVMSALNLSLHYNNKSTKSLFKPTLISLRLPTELRKELTEKLDQOIPR 85
 DB 160 HTKIFAGSGCYKMGITGEIESGYRIDLTRETSDDANNLKEGLIELDSNPVYV 219

QY 86 RPTHTVNTTISATHTATILNPRDTCRGDQ---HLLLEVRLHGR---KQYGGFL 138
 DB 220 MPTWGNDDTNPDSIAQVI-AELKYLKEFLSKYIILKHVPLLYKAKANKKELSGVLI 278

QY 139 RARMSPALMAGASGKVTDFNNNGTLYVSPFLPMEQGVSLILLHPSEGSALMSRRNG 198
 DB 279 DVAMDANLVLAATDILITDFS-----SVFEDVLTDKRITFY-----AMDDIYS 323

QY 199 YDKVIFTGQFVNGTSGVHSECGLLNTNAELCOYLDNRDQEGFYCYRPHMPCALTHMY 258
 DB 324 EDR---GMYLD-MEELPAP---ILKTVIELADYLSDDO-----L 356

QY 259 SKNKVSYLSKQEKSL-FERSNVGVEIM-----EKFNITSVSKCNTLKSVDLHESGKLQ 311
 DB 357 SDYLGKYLKAKEKYVYPDQGNSEIRIVQIFRKKESSQLVYKKEKLLFTFGMID 416

QY 312 HOLAVDLDRNIN--IOMOKICYPLI-----GSMTSVKEMEYLTTRAIDRTG---GEKNT 360
 DB 417 NNGITOSFINTLNALDYOKYIVVETNTPKSHFPHNYQKLKKNIRLIFRTGSPNSESQ 476

QY 361 VI--VISTDQHRPPPIIDVFIIRA-----LNVKRAIOHLLRSPDTMWIKTENIREMY 412
 DB 427 MLIHKINKSGHITSLPEVAFKREAHRLFGSLSHKAI----- 513

QY 413 NDAERSDHFHYOYLIIKIDFODLSYII 442
 DB 514 -----DFSGSYTWKSFVAFSDSRVKMI 536

RESULT 5
 T13709
 hypothetical protein F58E2.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 08-Sep-2000

C:Accession: T13709

K. Bolotin, D. Telehanuty, A.

submitted to the EMBL Data Library, October 1998

A:Description: The sequence of C. elegans cosmid F58E2.

A:Reference number: T21490

A:Accession: T13709

A:Accession: T13709

A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-688 <GCE>
 A:Cross-references: EMBL:AF100659; PIDN:AAK068966.1; GSPDB:GN00022; CESP:F58E2.3
 A:Experimental source: strain Bristol NZ; clone F58E2
 C:Genetics:
 A:Gene: CESP:F58E2.3
 A:Map position: 4
 A:Introns: 227/3; 306/3; 341/2; 602/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein F58E2.3

Query Match 4.4%; Score 109; DB 2; Length 688;
 Best Local Similarity 19.3%; Pred. No. 2.3;
 Matches 112; Conservative 76; Mismatches 200; Indels 192; Gaps 27;

QY 1 WKISINVKSLALLFLIASMTITVQNSTKYMSNLISLHY-----NSTKSLF 54
 DB 93 LEISQDIEEL-----DKMKSASQSTQFDWFOFLNNHRTGGLK 133

QY 55 PXTPLISL-----KPLTELRKELTEKLDQOIPR 86
 DB 134 PK--LISLFGTGYQALSFIRHISNNLEIINLNPREDTIEEHFOETVKK--KQKHA 169

QY 87 PPTHTVNTTISATHTATILNPRDTCRGDQHLHLLVROHLGRKQYGGFLAARSS 144
 DB 190 KYLHDEECTSISNTVHFGANFVVDYISIDLOLMEKLSHTSATLYKFNISF----- 242

QY 145 PALMAGASGKVTDFNNNGTLYVSPFLPMEQGVSLILLHPSEGSALMSRRNG 198
 DB 243 ---NNMADSDITDESITGLFGSPKHSLSFMEWSIDGIECSFEPFKFSIYCPKACEE 298

QY 194 ARNOGYDRVIFTGQFVNGTSGVHSECGLLNTNAELCOYLDNRDQEGFYCYRPHMPCALTHMY 258
 DB 299 LRN-----VQSKLEECTSLCTSELSIKYLNGLMKNILSHLGRDCLRL--RTSRG 350

QY 241 FYCYRPHMPCALTHMYSKNKVSYLSKQESLERSNVGVEYMEKNTISVSC----- 296
 DB 351 V-----QKVDIT-----KPAETIAEINLNVENVSVI-----FMAKIKKATYEP 394

QY 297 -----NTLKSVDLHESGKLOHOLAVDLDRNINIQM-----KYCPLIGSMT 338
 DB 395 IDLTTSYODHVSQVIOKVEKRNGLVERELC-----SVNNMQGSLSKFADLKLMI 449

QY 339 YSVKEME-----YLTRAIDRTGGEKNTYVI-----SLGHRFPPTIDVFI 379
 DB 450 AKKSEIQCVALDFSLTKANSRPGSKNKEHLLRFLAVAOGLDVISITDQDLOLQIPVKLS 509

QY 380 RRAINVKRAIOHLLRSPDTMWIKTENIREMYNDAREFSDPHGYOYLIIKDIFODLSV 439
 DB 510 IOVNLQETLLOLISLIRPSS---LKTIEIKKLND-----IISKSLKLNKA 553

QY 440 SIIDAW-----DITIAVGTNNVHPPO-----HYVQMOIIL 470
 DB 554 AQDDQMKNMEDILIEVL---ARQIOYQFLHVNTEINLM 590

RESULT 6
 LPHUB
 apolipoprotein B-100 precursor - human
 N:Contains: apolipoprotein B-26; apolipoprotein B-48; apolipoprotein B-74
 C:Species: Homo sapiens (man)
 C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 21-Jul-2000
 C:Accession: A27850; A25679; A25263; A25266; A24320; A24684; A23617; A25774
 A452; I61999; I59510; I39474; I39469; I84624; I31719; P50008
 R. Linder, E.H.; Blackhart, B.D.; Pletot, V.R.; Calati, L.; Fortier, C.; Knott, T.
 DNA 6, 363-372, 1987
 A:Title: DNA sequence of the human apolipoprotein B gene.
 A:Reference number: A27850; MUID:88003974
 A:Accession: A27850
 A:Molecule type: DNA
 A:Residues: 1-617 'A', 619-1929, 'P', 1931-3318, 'D', 3320-3426, 'T', 3428-3431, 'Q', 3433-3
 A:Cross-references: GB:M14162

R:Cladarias, C.; Hadzopoulou-Cladarias, M.; Nolte, R.T.; Atkinson, D.; Zannis, V.I.
EMBO J. 5, 3495-3507, 1986
A:Title: The complete sequence and structural analysis of human apolipoprotein B-100: re
A:Reference number: A91058; MUID:87161758
A:Accession: A25679
A:Molecule type: mRNA
A:Residues: 1-11, 15-2539, 'S', 2541-3823, 'R', 3825-4563 <CLA>
A>Note: 1109-Asp was also found
R:Knott, T.J.; Wallis, S.C.; Powell, L.M.; Pease, R.J.; Lusis, A.J.; Blackhart, B.; McCa
Nucleic Acids Res. 14, 7501-7503, 1986
A:Title: Complete cDNA and derived protein sequence of human apolipoprotein B-100.
A:Reference number: A93639; MUID:87016385
A:Accession: A25263
A:Molecule type: mRNA
A:Residues: 1-272, 'N', 274-617, 'A', 619-1217, 'E', 1219-2091, 'V', 2093-2364, 'T', 2366-2679, 'Q'
A:Cross-references: GB:X04506; NID:934330; PIDN:CAA28191.1; PID:934331
R:Law, S.W.; Grant, S.M.; Hlouch, K.; Hospattankar, A.; Lackner, K.; Lee, N.; Brewer JH
Proc. Natl. Acad. Sci. U.S.A. 83, 8142-8146, 1986
A:Title: Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino
A:Reference number: A94134; MUID:87041416
A:Accession: A25267
A:Molecule type: mRNA
A:Residues: 1-617, 'A', 619-703, 'P', 705-792, 'R', 794-1270, 'S', 1272-1866, 'G', 1868-2036, 'N', 2
4189-4220, 'W', 4222-4563 <LAW>
A>Note: the codons given for residues 704, 793, 1271, 2037, 2933, 3286, 3782, 4189, and
R:Chen, S.H.; Yang, C.Y.; Chen, P.F.; Setzer, D.; Tanimura, M.; Li, W.H.; Goto Jr., A.H
J. Biol. Chem. 261, 12918-12921, 1986
A:Title: The complete cDNA and amino acid sequence of human apolipoprotein B-100.
A:Reference number: A92556; MUID:87008488
A:Accession: A25266
A:Molecule type: mRNA
A:Residues: 1-97, 'T', 99-328, 'V', 330-644, 'T', 646-918, 'P', 920-3318, 'D', 3320-3426, 'T', 3428-
9-4137, 'G', 4134-4180, 'E', 4182-4563 <CHEN>
A:Cross-references: GB:J02610; NID:9178803; PIDN:AAA5549.1; PID:9178804
A>Note: a total of 2366 residues were confirmed by direct sequencing of tryptic peptides
R:Protter, A.A.; Hardman, D.A.; Sato, K.Y.; Schilling, J.W.; Yamanka, M.; Hoti, Y.D.; H
Proc. Natl. Acad. Sci. U.S.A. 83, 5678-5682, 1986
A:Title: Analysis of cDNA clones encoding the entire B-26 region of human apolipoprotein
A:Reference number: A24320; MUID:86287319
A:Accession: A24320
A:Molecule type: mRNA
A:Residues: 1-97, 'T', 99-617, 'A', 619-941, 'Y', 943-1138, 'P', 1140-178792
A:Cross-references: GB:M14081; NID:9178795; PIDN:AAA51752.1; PID:9553189
R:Law, S.W.; Lackner, K.J.; Hospattankar, A.V.; Anchors, J.M.; Sakaguchi, A.Y.; Naylor,
Proc. Natl. Acad. Sci. U.S.A. 82, 8340-8344, 1985
A:Title: Human apolipoprotein B-100: cloning, analysis of liver mRNA, and assignment of
A:Reference number: A24684; MUID:86094221
A:Accession: A24684
A:Molecule type: mRNA
A:Residues: 485-617, 'A', 619-1044 <LA2>
A:Cross-references: GB:M12480; NID:9178791; PIDN:AAA51751.1; PID:9178792
R:Protter, A.A.; Hardman, D.A.; Schilling, J.W.; Miller, J.; Appleby, V.; Chen, G.C.; K
Proc. Natl. Acad. Sci. U.S.A. 83, 1467-1471, 1986
A:Title: Isolation of a cDNA clone encoding the amino-terminal region of human apolipop
A:Reference number: A94088; MUID:86149925
A:Accession: A23817
A:Molecule type: mRNA
A:Residues: 1-291 <PRO>
A:Cross-references: GB:M12681; NID:9178797; PIDN:AAA51753.1; PID:9178798
R:Deeb, S.S.; Moczulsky, A.G.; Albers, J.J.
Proc. Natl. Acad. Sci. U.S.A. 82, 4983-4986, 1985
A:Title: A partial cDNA clone for human apolipoprotein B.
A:Reference number: A25774; MUID:85270450
A:Accession: A25774
A:Molecule type: mRNA
A:Residues: 709-791, 'SSSMKASHGCPHSGV', 810-906 <DEE>
A:Cross-references: GB:K03175; NID:9178821; PIDN:AAA51759.1; PID:9178822
R:Carlsson, P.; Danförs, C.; Olsson, S.O.; Bjursell, G.
Gene 49, 29-51, 1986
A:Title: Analysis of the human apolipoprotein B gene: complete structure of the B-74 reg
A:Reference number: A91565; MUID:87191999
A:Accession: A26533
A:Molecule type: mRNA
A:Residues: 1282-2711, 2742-3290, 'V', 3292-3336, 'N', 3338-3348, 'E', 3350-3361, 'V', 3363-3374, 'V', 3376-3387, 'V', 3389-3400, 'V', 3402-3413, 'V', 3415-3426, 'V', 3428-3439, 'V', 3441-3452, 'V', 3454-3465, 'V', 3467-3478, 'V', 3480-3491, 'V', 3493-3504, 'V', 3506-3517, 'V', 3519-3530, 'V', 3532-3543, 'V', 3545-3556, 'V', 3558-3569, 'V', 3571-3582, 'V', 3584-3595, 'V', 3597-3608, 'V', 3610-3621, 'V', 3623-3634, 'V', 3636-3647, 'V', 3649-3660, 'V', 3662-3673, 'V', 3675-3686, 'V', 3688-3699, 'V', 3701-3712, 'V', 3714-3725, 'V', 3727-3738, 'V', 3740-3751, 'V', 3753-3764, 'V', 3766-3777, 'V', 3779-3790, 'V', 3792-3803, 'V', 3805-3816, 'V', 3818-3829, 'V', 3831-3842, 'V', 3844-3855, 'V', 3857-3868, 'V', 3870-3881, 'V', 3883-3894, 'V', 3896-3907, 'V', 3909-3920, 'V', 3922-3933, 'V', 3935-3946, 'V', 3948-3959, 'V', 3961-3972, 'V', 3974-3985, 'V', 3987-3998, 'V', 4000-4011, 'V', 4013-4024, 'V', 4026-4037, 'V', 4039-4050, 'V', 4052-4063, 'V', 4065-4076, 'V', 4078-4089, 'V', 4091-4102, 'V', 4104-4115, 'V', 4117-4128, 'V', 4130-4141, 'V', 4143-4154, 'V', 4156-4167, 'V', 4169-4180, 'V', 4182-4193, 'V', 4195-4206, 'V', 4208-4219, 'V', 4221-4232, 'V', 4234-4245, 'V', 4247-4258, 'V', 4260-4271, 'V', 4273-4284, 'V', 4286-4297, 'V', 4299-4310, 'V', 4312-4323, 'V', 4325-4336, 'V', 4338-4349, 'V', 4351-4362, 'V', 4364-4375, 'V', 4377-4388, 'V', 4390-4401, 'V', 4403-4414, 'V', 4416-4427, 'V', 4429-4440, 'V', 4442-4453, 'V', 4455-4466, 'V', 4468-4479, 'V', 4481-4492, 'V', 4494-4505, 'V', 4507-4518, 'V', 4520-4531, 'V', 4533-4544, 'V', 4546-4557, 'V', 4559-4570, 'V', 4572-4583, 'V', 4585-4596, 'V', 4598-4609, 'V', 4611-4622, 'V', 4624-4635, 'V', 4637-4648, 'V', 4650-4661, 'V', 4663-4674, 'V', 4676-4687, 'V', 4689-4700, 'V', 4702-4713, 'V', 4715-4726, 'V', 4728-4739, 'V', 4741-4752, 'V', 4754-4765, 'V', 4767-4778, 'V', 4780-4791, 'V', 4793-4804, 'V', 4806-4817, 'V', 4819-4830, 'V', 4832-4843, 'V', 4845-4856, 'V', 4858-4869, 'V', 4871-4882, 'V', 4884-4895, 'V', 4897-4908, 'V', 4910-4921, 'V', 4923-4934, 'V', 4936-4947, 'V', 4949-4960, 'V', 4962-4973, 'V', 4975-4986, 'V', 4988-4999, 'V', 5001-5012, 'V', 5014-5025, 'V', 5027-5038, 'V', 5040-5051, 'V', 5053-5064, 'V', 5066-5077, 'V', 5079-5090, 'V', 5092-5103, 'V', 5105-5116, 'V', 5118-5129, 'V', 5131-5142, 'V', 5144-5155, 'V', 5157-5168, 'V', 5170-5181, 'V', 5183-5194, 'V', 5196-5207, 'V', 5209-5220, 'V', 5222-5233, 'V', 5235-5246, 'V', 5248-5259, 'V', 5261-5272, 'V', 5274-5285, 'V', 5287-5298, 'V', 5299-5310, 'V', 5312-5323, 'V', 5325-5336, 'V', 5338-5349, 'V', 5351-5362, 'V', 5364-5375, 'V', 5377-5388, 'V', 5390-5401, 'V', 5403-5414, 'V', 5416-5427, 'V', 5429-5440, 'V', 5442-5453, 'V', 5455-5466, 'V', 5468-5479, 'V', 5481-5492, 'V', 5494-5505, 'V', 5507-5518, 'V', 5520-5531, 'V', 5533-5544, 'V', 5546-5557, 'V', 5559-5570, 'V', 5572-5583, 'V', 5585-5596, 'V', 5598-5609, 'V', 5611-5622, 'V', 5624-5635, 'V', 5637-5648, 'V', 5650-5661, 'V', 5663-5674, 'V', 5676-5687, 'V', 5689-5700, 'V', 5702-5713, 'V', 5715-5726, 'V', 5728-5739, 'V', 5741-5752, 'V', 5754-5765, 'V', 5767-5778, 'V', 5780-5791, 'V', 5793-5804, 'V', 5806-5817, 'V', 5819-5830, 'V', 5832-5843, 'V', 5845-5856, 'V', 5858-5869, 'V', 5871-5882, 'V', 5884-5895, 'V', 5897-5908, 'V', 5910-5921, 'V', 5923-5934, 'V', 5936-5947, 'V', 5949-5960, 'V', 5962-5973, 'V', 5975-5986, 'V', 5988-5999, 'V', 6001-6012, 'V', 6014-6025, 'V', 6027-6038, 'V', 6040-6051, 'V', 6053-6064, 'V', 6066-6077, 'V', 6079-6090, 'V', 6092-6103, 'V', 6105-6116, 'V', 6118-6129, 'V', 6131-6142, 'V', 6144-6155, 'V', 6157-6168, 'V', 6170-6181, 'V', 6183-6194, 'V', 6196-6207, 'V', 6209-6220, 'V', 6222-6233, 'V', 6235-6246, 'V', 6248-6259, 'V', 6261-6272, 'V', 6274-6285, 'V', 6287-6298, 'V', 6299-6310, 'V', 6312-6323, 'V', 6325-6336, 'V', 6338-6349, 'V', 6351-6362, 'V', 6364-6375, 'V', 6377-6388, 'V', 6390-6401, 'V', 6403-6414, 'V', 6416-6427, 'V', 6429-6440, 'V', 6442-6453, 'V', 6455-6466, 'V', 6468-6479, 'V', 6481-6492, 'V', 6494-6505, 'V', 6507-6518, 'V', 6520-6531, 'V', 6533-6544, 'V', 6546-6557, 'V', 6559-6570, 'V', 6572-6583, 'V', 6585-6596, 'V', 6598-6609, 'V', 6611-6622, 'V', 6624-6635, 'V', 6637-6648, 'V', 6650-6661, 'V', 6663-6674, 'V', 6676-6687, 'V', 6689-6700, 'V', 6702-6713, 'V', 6715-6726, 'V', 6728-6739, 'V', 6741-6752, 'V', 6754-6765, 'V', 6767-6778, 'V', 6780-6791, 'V', 6793-6804, 'V', 6806-6817, 'V', 6819-6830, 'V', 6832-6843, 'V', 6845-6856, 'V', 6858-6869, 'V', 6871-6882, 'V', 6884-6895, 'V', 6897-6908, 'V', 6910-6921, 'V', 6923-6934, 'V', 6936-6947, 'V', 6949-6960, 'V', 6962-6973, 'V', 6975-6986, 'V', 6988-6999, 'V', 7001-7012, 'V', 7014-7025, 'V', 7027-7038, 'V', 7040-7051, 'V', 7053-7064, 'V', 7066-7077, 'V', 7079-7090, 'V', 7092-7103, 'V', 7105-7116, 'V', 7118-7129, 'V', 7131-7142, 'V', 7144-7155, 'V', 7157-7168, 'V', 7170-7181, 'V', 7183-7194, 'V', 7196-7207, 'V', 7209-7220, 'V', 7222-7233, 'V', 7235-7246, 'V', 7248-7259, 'V', 7261-7272, 'V', 7274-7285, 'V', 7287-7298, 'V', 7299-7310, 'V', 7312-7323, 'V', 7325-7336, 'V', 7338-7349, 'V', 7351-7362, 'V', 7364-7375, 'V', 7377-7388, 'V', 7390-7401, 'V', 7403-7414, 'V', 7416-7427, 'V', 7429-7440, 'V', 7442-7453, 'V', 7455-7466, 'V', 7468-7479, 'V', 7481-7492, 'V', 7494-7505, 'V', 7507-7518, 'V', 7520-7531, 'V', 7533-7544, 'V', 7546-7557, 'V', 7559-7570, 'V', 7572-7583, 'V', 7585-7596, 'V', 7598-7609, 'V', 7611-7622, 'V', 7624-7635, 'V', 7637-7648, 'V', 7650-7661, 'V', 7663-7674, 'V', 7676-7687, 'V', 7689-7700, 'V', 7702-7713, 'V', 7715-7726, 'V', 7728-7739, 'V', 7741-7752, 'V', 7754-7765, 'V', 7767-7778, 'V', 7780-7791, 'V', 7793-7804, 'V', 7806-7817, 'V', 7819-7830, 'V', 7832-7843, 'V', 7845-7856, 'V', 7858-7869, 'V', 7871-7882, 'V', 7884-7895, 'V', 7897-7908, 'V', 7910-7921, 'V', 7923-7934, 'V', 7936-7947, 'V', 7949-7960, 'V', 7962-7973, 'V', 7975-7986, 'V', 7988-7999, 'V', 8001-8012, 'V', 8014-8025, 'V', 8027-8038, 'V', 8040-8051, 'V', 8053-8064, 'V', 8066-8077, 'V', 8079-8090, 'V', 8092-8103, 'V', 8105-8116, 'V', 8118-8129, 'V', 8131-8142, 'V', 8144-8155, 'V', 8157-8168, 'V', 8170-8181, 'V', 8183-8194, 'V', 8196-8207, 'V', 8209-8220, 'V', 8222-8233, 'V', 8235-8246, 'V', 8248-8259, 'V', 8261-8272, 'V', 8274-8285, 'V', 8287-8298, 'V', 8299-8310, 'V', 8312-8323, 'V', 8325-8336, 'V', 8338-8349, 'V', 8351-8362, 'V', 8364-8375, 'V', 8377-8388, 'V', 8390-8401, 'V', 8403-8414, 'V', 8416-8427, 'V', 8429-8440, 'V', 8442-8453, 'V', 8455-8466, 'V', 8468-8479, 'V', 8481-8492, 'V', 8494-8505, 'V', 8507-8518, 'V', 8520-8531, 'V', 8533-8544, 'V', 8546-8557, 'V', 8559-8570, 'V', 8572-8583, 'V', 8585-8596, 'V', 8598-8609, 'V', 8611-8622, 'V', 8624-8635, 'V', 8637-8648, 'V', 8650-8661, 'V', 8663-8674, 'V', 8676-8687, 'V', 8689-8700, 'V', 8702-8713, 'V', 8715-8726, 'V', 8728-8739, 'V', 8741-8752, 'V', 8754-8765, 'V', 8767-8778, 'V', 8780-8791, 'V', 8793-8804, 'V', 8806-8817, 'V', 8819-8830, 'V', 8832-8843, 'V', 8845-8856, 'V', 8858-8869, 'V', 8871-8882, 'V', 8884-8895, 'V', 8897-8908, 'V', 8910-8921, 'V', 8923-8934, 'V', 8936-8947, 'V', 8949-8960, 'V', 8962-8973, 'V', 8975-8986, 'V', 8988-8999, 'V', 9001-9012, 'V', 9014-9025, 'V', 9027-9038, 'V', 9040-9051, 'V', 9053-9064, 'V', 9066-9077, 'V', 9079-9090, 'V', 9092-9103, 'V', 9105-9116, 'V', 9118-9129, 'V', 9131-9142, 'V', 9144-9155, 'V', 9157-9168, 'V', 9170-9181, 'V', 9183-9194, 'V', 9196-9207, 'V', 9209-9220, 'V', 9222-9233, 'V', 9235-9246, 'V', 9248-9259, 'V', 9261-9272, 'V', 9274-9285, 'V', 9287-9298, 'V', 9299-9310, 'V', 9312-9323, 'V', 9325-9336, 'V', 9338-9349, 'V', 9351-9362, 'V', 9364-9375, 'V', 9377-9388, 'V', 9390-9401, 'V', 9403-9414, 'V', 9416-9427, 'V', 9429-9440, 'V', 9442-9453, 'V', 9455-9466, 'V', 9468-9479, 'V', 9481-9492, 'V', 9494-9505, 'V', 9507-9518, 'V', 9520-9531, 'V', 9533-9544, 'V', 9546-9557, 'V', 9559-9570, 'V', 9572-9583, 'V', 9585-9596, 'V', 9598-9609, 'V', 9611-9622, 'V', 9624-9635, 'V', 9637-9648, 'V', 9650-9661, 'V', 9663-9674, 'V', 9676-9687, 'V', 9689-9700, 'V', 9702-9713, 'V', 9715-9726, 'V', 9728-9739, 'V', 9741-9752, 'V', 9754-9765, 'V', 9767-9778, 'V', 9780-9791, 'V', 9793-9804, 'V', 9806-9817, 'V', 9819-9830, 'V', 9832-9843, 'V', 9845-9856, 'V', 9858-9869, 'V', 9871-9882, 'V', 9884-9895, 'V', 9897-9908, 'V', 9910-9921, 'V', 9923-9934, 'V', 9936-9947, 'V', 9949-9960, 'V', 9962-9973, 'V', 9975-9986, 'V', 9988-9999, 'V', 10001-10012, 'V', 10014-10025, 'V', 10027-10038, 'V', 10040-10051, 'V', 10053-10064, 'V', 10066-10077, 'V', 10079-10090, 'V', 10092-10103, 'V', 10105-10116, 'V', 10118-10129, 'V', 10131-10142, 'V', 10144-10155, 'V', 10157-10168, 'V', 10170-10181, 'V', 10183-10194, 'V', 10196-10207, 'V', 10209-10220, 'V', 10222-10233, 'V', 10235-10246, 'V', 10248-10259, 'V', 10261-10272, 'V', 10274-10285, 'V', 10287-10298, 'V', 10299-10310, 'V', 10312-10323, 'V', 10325-10336, 'V', 10338-10349, 'V', 10351-10362, 'V', 10364-10375, 'V', 10377-10388, 'V', 10390-10401, 'V', 10403-10414, 'V', 10416-10427, 'V', 10429-10440, 'V', 10442-10453, 'V', 10455-10466, 'V', 10468-10479, 'V', 10481-10492, 'V', 10494-10505, 'V', 10507-10518, 'V', 10520-10531, 'V', 10533-10544, 'V', 10546-10557, 'V', 10559-10570, 'V', 10572-10583, 'V', 10585-10596, 'V', 10598-10609, 'V', 10611-10622, 'V', 10624-10635, 'V', 10637-10648, 'V', 10650-10661, 'V', 10663-10674, 'V', 10676-10687, 'V', 10689-10700, 'V', 10702-10713, 'V', 10715-10726, 'V', 10728-10739, 'V', 10741-10752, 'V', 10754-10765, 'V', 10767-10778, 'V', 10780-10791, 'V', 10793-10804, 'V', 10806-10817, 'V', 10819-10830, 'V', 10832-10843, 'V', 10845-10856, 'V', 10858-10869, 'V', 10871-10882, 'V', 10884-10895, 'V', 10897-10908, 'V', 10910-10921, 'V', 10923-10934, 'V', 10936-10947, 'V', 10949-10960, 'V', 10962-10973, 'V', 10975-10986, 'V', 10988-10999, 'V', 11001-11012, 'V', 11014-11025, 'V', 11027-11038, 'V', 11040-11051, 'V', 11053-11064, 'V', 11066-11077, 'V', 11079-11090, 'V', 11092-11103, 'V', 11105-11116, 'V', 11118-11129, 'V', 11131-11142, 'V', 11144-11155, 'V', 11157-11168, 'V', 11170-11181, 'V', 11183-11194, 'V', 11196-11207, 'V', 11209-11220, 'V', 11222-11233, 'V', 11235-11246, 'V', 11248-11259, 'V', 11261-11272, 'V', 11274-11285, 'V', 11287-11298, 'V', 11299-11310, 'V', 11312-11323, 'V', 11325-11336, 'V', 11338-11349, 'V', 11351-11362, 'V', 11364-11375, 'V', 11377-11388, 'V', 11390-11401, 'V', 11403-11414, 'V', 11416-11427, 'V', 11429-11440, 'V', 11442-11453, 'V', 11455-11466, 'V', 11468-11479, 'V', 11481-11492, 'V', 11494-11505, 'V', 11507-11518, 'V', 11520-11531, 'V', 11533-11544, 'V', 11546-11557, 'V', 11559-11570, 'V', 11572-11583, 'V', 11585-11596, 'V', 11598-11609, 'V', 11611-11622, 'V', 11624-11635, 'V', 11637-11648, 'V', 11650-11661, 'V', 11663-11674, 'V', 11676-11687, 'V', 11689-11700, 'V', 11702-11713, 'V', 11715-11726, 'V', 11728-11739, 'V', 11741-11752, 'V', 11754-11765, 'V', 11767-11778, 'V', 11780-11791, 'V', 11793-11804, 'V', 11806-11817, 'V', 11819-11830, 'V', 11832-11843, 'V', 11845-11856, 'V', 11858-11869, 'V', 11871-11882, 'V', 11884-11895, 'V', 11897-11908, 'V', 11910-11921, 'V', 11923-11934, 'V', 11936-11947, 'V', 11949-11960, 'V', 11962-11973, 'V', 11975-11986, 'V', 11988-11999, 'V', 12001-12012, 'V', 12014-12025, 'V', 12027-12038, 'V', 12040-12051, 'V', 12053-12064, 'V', 12066-12077, 'V', 12079-12090, 'V', 12092-12103, 'V', 12105-12116, 'V', 12118-12129, 'V', 12131-12142, 'V', 12144-12155, 'V', 12157-12168, 'V', 12170-12181, 'V', 12183-12194, 'V', 12196-12207, 'V', 12209-12220, 'V', 12222-

ch encodes the 250k apob-48. CAA encoding 2180-gln is substituted by the stop codon TAA.
 PrYang, C.; Kim, T.M.; Wang, S.; Lee, B.; Yang, M.; Gotto Jr., A.M.
 Proc. Natl. Acad. Sci. U.S.A. 87, 5523-5527, 1990
 A>Title: Isolation and characterization of sulfhydryl and disulfide peptides of human ap
 A:Reference number: A35783; M0ID:90319144
 A:Contents: disulfide bonds
 A:Accession: A35783
 A:Molecule type: protein
 A:Residues: 28-41;76-97, 'T', 99-100;175-193;206-215;239-249;259-266;357-399;455-490;512-5
 A>Note: Cysteines at positions 112, 1422, 1505, 1662, 3761, 3917, and 4217 have free su
 R:Lehoucq, R.C.; Miller, C.; Shively, J.E.; Schumaker, V.N.; Balla, M.A.; Lusis, A.J.
 FEBS Lett. 170, 105-108, 1984
 A>Title: Human apolipoprotein B: partial amino acid sequence.
 A:Reference number: A22006; M0ID:84208786
 A:Accession: A22006
 A:Molecule type: protein
 A:Residues: 873-892, 'K', 894-896 <LE1>
 A:Accession: h22006
 A:Molecule type: protein
 A:Residues: 3113, 'L', 3115-3130, 'R', 3132-3133, 'P', 3135-3136, 'R' <LE2>
 R:Blackhart, B.D.; Ludwig, E.M.; Pierotti, V.R.; Calafel, L.; Onasch, M.A.; Wallis, S.C.;
 J. Biol. Chem. 261, 15364-15367, 1986
 A>Title: Structure of the human apolipoprotein B gene.
 A:Reference number: A92564; M0ID:87057153
 A:Contents: annotation; gene structure
 R:Wagener, R.; Pitzner, R.; Stoffel, W.
 Biol. Chem. Hoppe-Seyler 368, 419-425, 1987
 A>Title: Studies on the organization of the human apolipoprotein B 100 gene.
 A:Reference number: A90715; M0ID:87271140
 A:Contents: annotation; gene structure
 R:Weisgraber, K.H.; Rall Jr., S.C.
 J. Biol. Chem. 262, 11097-11103, 1987
 A>Title: Human apolipoprotein B-100 heparin-binding sites.
 A:Reference number: A92605; M0ID:87280197
 A:Contents: annotation; heparin binding and disulfide bond
 R:Usahil, N.; Lee, D.M.; Mok, T.
 Biochem. Biophys. Res. Commun. 137, 493-499, 1986
 A>Title: Apolipoprotein B is a calcium binding protein.
 A:Reference number: A90125; M0ID:86242245
 A:Contents: annotation; calcium binding
 R:Carlsson, P.; Olofsson, S.O.; Bondieters, G.; Darifors, C.; Wiklund, O.; Blursell, G.
 Nucl. Acids Res. 13, 8813-8826, 1985
 A>Title: Molecular cloning of human apolipoprotein B cDNA.
 A:Reference number: 137178; M0ID:86093680
 A:Accession: 137180

Query Match 4.2%; Score 104.5; DB 1; Length 4563;
 Best local similarity 20.5%; Pred. No. 75;
 Matches 89; Conservative 72; Mismatches 169; Indels 105; Gaps 22;

56 KTPILSLKPLETE-LRIKELIKLDOOIPRPETHV---NTTSATHS-TATILNPD 109
 2025 KVPLLSEPIINIDALEMRDAVEK-----POEFTVAEYKQDQVSHINLPFEETQ 2078
 110 TYCRDQDLHILEVRDHLGRKQYGD-FLRARRSPALMAGASGYVDENGYIVSYFT 168
 2079 EYFERNRQTIIVLENVQNRKHNINIDQFAK-----YRAALKLPQOAN-DYLSNRN 2130
 169 LFMEGQVSLSLILHSEVSAWS-----AANQGYRVTF 204
 2141 --WERQVS-----HAKERKLTALTKKRYTENDTQALDANKINENKLSQLOQY--MIO 2180
 205 TQGFVNGTSOVHSEGLIINTNAELQYLDNRDQEGFYCVRPQHPFCALTIHWYSK---- 260
 2181 PDQYIKDSYDLHDLKIAIANITIDEIKLSDEH--YHIRNLYKTIHDLHLFINIDF 2238
 261 NKKVSYLSKQKSLFPRSNVGVETMEKFNITISVSKNTKLSVDL-HESGKL-OHQLAVDL 318
 2239 NKSQSTASWIDQVTRKQYRIQIUEKIQDK---RHQNDIDQHLGKLGKQKHQIAIDV 2294
 419 DNNINIQWKQYCPILIGSMYTSVKEMEYLTRALDRTGGEKNTVVIYISLGHFRPPPI--- 375

Db 2295 -----RVLLDQLGT-----TISFERI-----NDVL-----EHKKHPIVNL 2325
 QY 376 -DVFIRALNWKALQIHLLRSPDT-----MYITKENREMYNDAREFSDFHGYIOL 429
 Db 2326 GDFEVAERKINARVAVHELLERYEDQOIVLMKLVELAHQYKKEITIKLSVNLQOVK 2385
 QY 430 IKDIFDLSVSIIDA 444
 Db 2386 IKQYFEKLVGPIIDA 2400

RESULT 7
 VBEB6
 glycoprotein H precursor - human herpesvirus 6 (strains U1102, GS, and AJ)
 C:Species: human herpesvirus 6
 C:Date: 31-Dec-1992 *sequence_revision 31-Dec-1992 *text_change 24-Sep-1999
 C:Accession: A40511; J02382; P00846; D56653; P00847
 R:Josephs, S.F.; Ablashi, D.V.; Salahuddin, S.Z.; Jagodzinski, L.L.; Wong-Staal, F.;
 J. Virol. 65, 5597-5604, 1991
 A>Title: Identification of the human herpesvirus 6 glycoprotein H and putative large
 A:Reference number: A40511; M0ID:91374623
 A:Accession: A40511
 A:Molecule type: DNA
 A:Residues: 1-694 <J08>
 A:Cross-references: GB:S57509; NID:9235428; PIDN:AAJ19779.1; PID:q235431
 A:Experimental source: strain GS, clone p2VB43
 R:Compeis, U.A.; Carigan, D.R.; Cariss, A.L.; Arno, J.
 J. Gen. Virol. 74, 613-622, 1993
 A>Title: Two groups of human herpesvirus 6 identified by sequence analyses of labora
 A:Reference number: J02382; M0ID:93224882
 A:Accession: J02382
 A:Molecule type: DNA
 A:Residues: 1-694 <GOM>
 A:Cross-references: GB:X83413; NID:9853961; PIDN:CAA58382.1; PID:9854027
 A:Experimental source: strain AJ
 A:Molecule type: DNA
 A:Residues: 1-65 <GOM1>
 A:Experimental source: strain GS
 R:Compeis, U.A.; Cariss, A.L.; Sun, N.; Arrand, J.R.
 DNA Seq. 3, 25-39, 1992
 A>Title: Infectivity determinants encoded in a conserved gene block of human herpes
 A:Reference number: A56653; M0ID:99091236
 A:Accession: D56653
 A:Molecule type: DNA
 A:Residues: 1-694 <G02>
 A:Cross-references: GB:X83413; NID:9853961; PIDN:CAA58382.1; PID:9854027
 A:Experimental source: strain U1102
 A>Note: sequence extracted from NCBI backbone (NCBIN:120533, NCBI:121866)
 C:Superfamily: herpesvirus glycoprotein H
 C:Keywords: glycoprotein; transmembrane protein
 F:1-23/Domain: signal sequence *status predicted <SIG>
 F:24-694/Product: glycoprotein H *status predicted <GPI>
 F:554-570/Domain: transmembrane *status predicted <TM2>
 F:668-684/Domain: transmembrane *status predicted <TM2>
 F:21,24,51,76,148,206,337,365,384,411,538,573,593,652/binding site: carbohydrate (As

Query Match 4.2%; Score 104; DB 1; Length 694;
 Best local similarity 19.4%; Pred. No. 5.6;
 Matches 101; Conservative 77; Mismatches 208; Indels 134; Gaps 24;

17 ILASWILITPVNSTKWSALNLSLSLHYNNSTKSLFKPKPLSLKPLETLEIKETI 76
 Db 2 LRLWVP--VLLPFCYGMKPRPLMISNSHCRNGFNENPIVRPQFTTFNFYTKNDTRIYQVP 59
 QY 77 E-KLDQOIPRPETHVNTTSATHTAIIIN-----PRDYCRGQDLHILEVRDHLGRK 131
 Db 60 KCLGSDITLHLDIAINTESLTNTEKRYRVEPEPMN----- 97
 QY 132 QYSGDFLRARRSSPALMAGASGYVDNRNNGTIVSYFTLFMEGQVSLSLILHPSGCV--S 189
 Db 98 -----DILRL-----SPV-----PSYKQFNIDRSI-----QPVVYSLSNM-YPSQGIYVV 136

```

Oy 190 ALMSARNOGYDRVIFGQFVNGTSQ-----VHSEGGILNTNNAELCOYLDRDDEGFCYR 245
Db 137 RVEEVRNOGYDNV---SCKLPNSLAKELFFVOYRC-----AKITRYGEDIYTHFF--T 185
Oy 246 PQHM-----PCALTHMSKKKKVSYLSKOES--LFRSNVNG-----VEIMEK 287
Db 186 PQHMILYQNPAGDLTMMYGNTTSINFAPYKKSSFFKQTLIDLLIYKENDIYQYR 245
Oy 288 FNTISVSKCNTLKSVDLHESGKLQHLAVDLDRNINIQ-----WKKCYPLGSM--TY 339
Db 246 FTSDFEVDLNDVDEVALLLKF-----NNLDIQLLRGDOCKRPNAGIPQMMELY 298
Oy 340 SVKEMEYLTIRADRTGGEKNT---VIYSLGQHFPPIDVFIRBALNVH-----KAI 389
Db 299 GIVHFSYST-----KNIGPMFLVRLKTHENTLSDSFVNCVNSEGLTQYPMK 349
Oy 390 QHLLRSPDTMVIKTEKNT-----REMYNDAERFSDPFGYTOYLLIKDIFQ 435
Db 350 EFLKXEPDSYITKKKSISVTLTYLATAYESNTISKYKMTDINTQNIYEKHM--F 408
Oy 436 DLVSIIIDAMDITVAGTNVHPQHVVGQIMILNITIC 475
Db 409 FNLTFSDRETLPMLEINIIPTDERMORHMLIGNDC 448

```

```

RESULT 8
S73779
N:leucine--tRNA ligase (EC 6.1.1.4) leus - Mycoplasma pneumoniae (strain ATCC 29342)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C>Date: 27-Feb-1997 #sequence-revision 25-Apr-1997 #text-change 07-Dec-1999
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkil, E.; Li, B. C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A>Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73779; MUID:97105885
A:Accession: S73779
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-793 <HMD>
A:Cross-references: EMBL:AE000044; GB:U00089; NID:g1674130; PIDN:AAB96101.1; PID:g167414
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Gene: leus
A:Genetic code: SGC3
C:Superfamily: leucine--tRNA ligase
C:Keywords: aminocyl-tRNA synthetase; ligase; protein biosynthesis

```

```

Query Match 4.1% Score 103; DB 2; Length 793;
Best Local Similarity 19.2% Pred. No. 8.1;
Matches 86; Conservative 85; Mismatches 174; Indels 102; Gaps 20;
Oy 1 MKISMINKSLALLFLASMIITFVONSTRKWSALNLSISLHYNNSTKSLFPPKPII 60
Db 306 LKTKALNPAINTAIPLYVANYVEGYTDAV-----MGVAHNENNPFAKAKLPIT 358
Oy 61 SLKPLTELRIKEIKEDQIPRPFTYNTTSTSTATILNPPDYCRGDDLHL 120
Db 359 T-----VIDKQERLQHSQPSGINSOTANTQITMLVERKA-----KKTIV 400
Oy 121 LEVRDHLRRKQYGGDFLRARSSPALMAGASGVTDFFNNGTYLVSFLPEWGVSLSL 180
Db 401 YKLRDMIFSQRQWGE-----PPIPLF-----DENNGHLVK-----ELPYTLRAL 441
Oy 181 LIHPSEGV--ALMSARNOGYDRVIFGQ--FVNGTS--QVHSEC-----GLIINTNAELC 230
Db 442 ANYPDQSTNPLM--RNOEMAKVQGNQITRETISTPMQWAGSCWYIYLMLINNEF 499
Oy 231 OYLDNRDEGFCYVRPQHMP-----AALTHMSKKKKVSYLSKOESL 274

```

```

Db 500 WPIDREAKDLF---ERYLPVDLYGGAENAVLHLLXAPMHPQIKKQIVTKREFKKL 546
Oy 275 FRSNVGVELMEFNFTISVSKCNTLKS---VDI HESGKLQHLAVDLDRNINIQ--KAY 341
Db 557 T-----NQGAVLQPPGKKSSKSGNINPTPLIDSHGADALRYL-----MHPG 603
Oy 332 PLIGSMYSVKEMEYLTIRADRTGQ--EKNIVVISLQHFPPIDVFIRBALNVHAF 409
Db 601 PLSAALTMWDGJNGRMRBLDRVINFFHKNENIKETVQDE--IVSYNMLFKRSFE--HLEK 659
Oy 390 QHLLRSPDTMVIKTEKNT-----KTEINREMY 412
Db 659 QELNLYISQMMIFLNLYKTKQTLILAY 685

```

```

RESULT 9
B44858
N:lactocoeptin (EC 3.4.21.96) precursor [validated] - Lactobacillus paracasei (strain N)
A:Alternate names: cell-envelope-associated proteinase prtP; serine proteinase; Lf
C:Species: Lactobacillus paracasei
C>Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 19-May-2000
C:Accession: B44858; C44858; A44850
R:Holck, A.; Naes, H.
J. Gen. Microbiol. 138, 1353-1364, 1992
A>Title: Cloning, sequencing and expression of the gene encoding the cell-envelope-associated proteinase prtP from Lactobacillus paracasei
A:Reference number: A44858; MUID:92381481
A:Accession: B44858
A:Molecule type: DNA
A:Residues: 1-1902 <HOL1>
A:Cross-references: GB:M83946; NID:g149580; PIDN:AAA5248.1; PID:g149582
A>Note: sequence extracted from NCBI backbone (NCBI:112261, NCBI:112263)
A:Accession: C44858
A:Molecule type: protein
A:Residues: X', 189-196 <HOL2>
R:Naes, H.; Nissen-Meyer, J.
J. Gen. Microbiol. 138, 313-318, 1992
A>Title: Purification and N-terminal amino acid sequence determination of the cell-envelope-associated proteinase prtP from Lactobacillus paracasei
A:Reference number: A44850; MUID:92226694
A:Accession: A44850
A:Status: preliminary
A:Molecule type: protein
A:Residues: 189-196 <NAE>
A:Cross-references: PIDN:AAB22052.1; PID:g248666
A:Experimental source: strain NCDO 151
A>Note: sequence extracted from NCBI backbone (NCBI:94706)
C:Genetics:
A:Gene: prtP
C:Superfamily: lactocoeptin; subtilisin homology
C:Keywords: hydrolase; serine proteinase; transmembrane protein
F:1-23/Domain: signal sequence; status predicted <Sig>
F:23-187/Domain: propeptide; status predicted <PKO>
F:188-1902/Product: serine proteinase; cell-envelope-associated; status experiment
F:208-634/Domain: subtilisin homology; status atypical <SBI>

```

```

Query Match 4.1% Score 101.5; DB 1; Length 1902;
Best Local Similarity 21.6% Pred. No. 37;
Matches 102; Conservative 47; Mismatches 159; Indels 165; Gaps 24;
Oy 58 PLISAKPLTELRIK-----EIEKLDQIPRPFTYNTTSTSTATILNPR 108
Db 714 PAVELKDEFTSDTFKLTFTNTTHETLYQMS-----NIDTNAVYTSATQPN-- 761
Oy 109 DTYCRGQDLHILLEVRDHLGRKQYGGDFLRARSSPALMAGASGVTDFFNNGTYLVSFL 168
Db 762 -----SGVLDKKIDDAIKASD--ITVAGKTAQLEFT 794
Oy 169 L-----FMQGVSLSLILHPSEGVSLMSARNOGYDRVIFGQFVNG--TSOVS 217
Db 795 LSLKSPDOQGFVEGFLNFKGSDG-----SRLLUPY--MGFFGDMNDKIVDSLNGITYS 847
Oy 218 ECG-----LILNTNAELCOYLDN--RDQEGFCYVRPQHMP--CALTHMSKKKKVSYLS 246

```


RESULT 12

T42631

breast cancer tumor suppressor BRCA2 - dog (fragment)

C:Species: Canis lupus familiaris (dog)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C:Accession: T42631

R:Biggell, G.R.; Micklem, G.; Stratton, M.R.; Ashworth, A.; Wooster, R.

A:Description: BRCA2 contains a novel repeat that is conserved in other mammalian species.

A:Reference number: 222184

A:Accession: T42631

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1507 <BIG>

A:Cross-references: EMBL:Z75664; NID:e996328; PIR:e251256; PIDD:CAA9994.1

C:Genetics:

A:Gene: BRCA2

C:Superfamily: breast cancer tumor suppressor BRCA2

C:Keywords: tumor suppressor

Query Match 4.0%; Score 99.5; DB 2; Length 1507;

Best Local Similarity 20.6%; Pred. No. 38;

Matches 77; Conservative 55; Mismatches 143; Indels 99; Gaps 18;

```

OY 22 IFTVONSTKWSALNLSISLHYNNSTKSLFPTPLSLK-----PLTELE- 70
   || ||||| |
DB 436 IISGFVQNSTYV-----SDSESCHTAPPTLSLKODSDSNKRLTPSQKKEIT 481
OY 71 RIKEITKIDQIIPRP---THV--NTTSATSTATILNPRDYCRGDLHLEVR- 124
   || ||||| |
DB 482 ELSTILESGSOFETQFRKPSHILOKNPFPMPENQTLITNSTKEMDODLHITNAPS 541
OY 125 -DHLGRKQ--YSGDELRAKSSPALMAGASGYVDNFNGTYLVSTFLMEGO-VSLSL 179
   || ||||| |
DB 542 ISQVDSKSKSEGIQKQKQFACISRTSCNRSASGYSTDKNE---VEFGTSARGLKLV 597
OY 180 LLIHPSEGSALMSAR-----NOGYDRVLTGQGVNGTSOV----- 215
   || ||||| |
DB 598 -----GSEALQAKKFLSDLENIETSYEVDRSSTSKNDSDVSMIQEDCNKLN 650
OY 216 --HSECGILNTNAELC-----OYLDNRDQEGFYCVRPHMPCALT--HMYSKN 261
   || ||||| |
DB 651 EPNNKCRLLIIONNIENTDIFVEEYTESYRNTENGNCCTAGANTCNSESDSSSKN 710
OY 262 KKVYLSKQKSL--FERSNGVEIMEKF-----NTISVKCNTLKSVDLHESKIQH--- 312
   || ||||| |
DB 711 DTV-YIHEENGLPICDQNNIDKLFSQFMKEGNTQIKESGLDLCLEVMKAEETSHVM 769
OY 313 -----QLAVLDLDRNI 322
   || ||||| |
DB 770 SNKQDLTANTGONI 783

```

RESULT 13

T07709

hypothetical protein T23J7.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 24-Nov-1999

C:Accession: T07709

R:Queller, F.; Choinsie, N.; Robert, C.; Brothier, P.; Wincker, P.; Catolico, L.; Artigiani

A:Reference number: 215793

A:Accession: T07709

A:Molecule type: DNA

A:Residues: 1-795 <QUE>

A:Cross-references: EMBL:AL049746; GSPDB:GN00061; ATSP:T23J7.30

A:Experimental source: cultivar Columbia; BAC clone T23J7

A:Gene: ATSP:T23J7.30

A:Map position: 3

A:Introns: 99/1; 133/1; 166/3

C:Superfamily: Arabidopsis thaliana hypothetical protein T23J7.30

Query Match

Best Local Similarity 4.0%; Score 98.5; DB 2; Length 795;

Matches 100; Conservative 79; Mismatches 181; Indels 103; Gaps 27;

```

OY 33 VMSALNLSISLHYNNSTKSLFPTPLSLKLTITE-----LRKEITKIDQI 82
   || ||||| |
DB 149 VMSLKNL-----RTSSSGFEVRLHAIKTKTEITSSVAKHHPMALYSAAVDIR 204
OY 83 I-----PPRFTVNTTSAT--HSTATILNPRDYCRGDLHLEVR- 124
   || ||||| |
DB 204 VDALAMPQAIADYRALLSLRMPQ-----LSLTSASLDSKSSNQN- 250
OY 117 LHILEVRDHLGRKQYGGDFLRAKSSPALMAGASGYVDNFNGTYLVSTFLMEGOVS 176
   || ||||| |
DB 251 --LFMEGSL--KSYCGSF-HALCSLOGLOLQKRSROLGIHKGENVLFQPLMAIEEL 304
OY 177 LSLILLHPSGVSAKARNQGYRVIF--TGQFVNGTSVHSEGLIINTNAELQYLD 214
   || ||||| |
DB 305 VNPPLIV-ASQRRFTKWEKPEFIALVYKTRDYDSMDL-----LQPLIV 350
OY 235 NRQDEGYCVRPHMPCALTHY-----SKKKKSYL----- 267
   || ||||| |
DB 351 EAKLAGYSCREWVSAMVSSLSLYLKEIPFIYVQGLDEANETDLRSKAKVSHLIDLM 410
OY 268 ---SKQKSLFERNVGVLEMEKFTISVK---CNTLSVDLHESKIQHACLA----- 315
   || ||||| |
DB 411 ISFDKRVSLVSQGI--LSLQEDGMLTSSLSYFCRPMQDLMAETELDERLYKKE 466
OY 316 VLDLNNINIQOK-----YCPILGSMYSVKEKMEYLFRAIDRTGKENVIVISG 367
   || ||||| |
DB 470 INDNRMTAKVQDELSSSNVYRPILISIF-----LQHLSSILERS---KVPALYLR 520
OY 368 QHFRFPIDVFIRALN--VHKAIDHLLRSPT--HYIKTINIEKMDAERPSDFHY 424
   || ||||| |
DB 521 AR-----FLRLASPTIHKFLDCLLRQDQDLTLITN-----NDLIVSNISNA 567
OY 425 IYLLIKIKIFQDLSSVSIIDAMDITAY---GTNNVHPRHVYG 464
   || ||||| |
DB 568 GHYI-----ESVLEWSEDFELEMGTQ--HDQEVHG 599

```

RESULT 14

A71617

SEBA antigen/papain-like proteinase with active Ser PFB0360C - malaria parasite (P. falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: A71617

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Atwood, L.; Kocun, S.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, K.; White, G.; Smith

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum

A:Reference number: A71600; M0UD:99021743

A:Accession: A71617

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-11004 <GAR>

A:Cross-references: GB:AE001389; GB:AE001362; NID:93845161; PIDD:AA071461.1; PIR:111111

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0360C

C:Superfamily: Plasmodium vivax serine-repeat antigen

Query Match

Best Local Similarity 4.0%; Score 98.5; DB 2; Length 1004;

Matches 85; Conservative 88; Mismatches 151; Indels 141; Gaps 22;

```

OY 23 IFTVONSTKWSALNLSISLHYNNSTKSLFPTPLSLKLTTELRKEITKIDQI 82
   || ||||| |

```

```

Db      71  HAASFSDTKEMSPNNLT-----SKKKKKKKELRPDMSNDS- 109
QY      83  IPRPFTHVNTTTSATHSTATILNPRDTCYRGQDLHLLVBDHLGRKQYGGDFLRAM 142
Db      110  -----SMTSSINKON-----NNOIKSVL-LKENKG-----VKI 136
QY      143  SSPALMAGASQKVTDFNNGTYLVSTFLWEGVSLILLI-HPSGVSAIAMSARNGCYR 201
Db      137  TGPC-----NNLSIFLVPIITYIDETKYNIELEKYL 169
QY      202  VLTG--QFVNGTSQVHSECGILLTNAELQYLDRDQEGFYCRQHPHCAALTHMYS 259
Db      170  DEFSISIKFKPDITTELRTSDDTLMNTNPNVGSRDPLDRLYNI-----CA- 216
QY      260  KKKKSVYLSKQESLPERSNVGEIEMKFNITSVSKCNTLKSVDLHESGLOQLAVDLD 319
Db      217  EKTFKFFV-----VYIKNT-----LTLK-WKYETGVTKNK--VDIR 251
QY      320  RNINIQOKYCYPLIGSMITSYKMEYLTRAIDRTGGEKNTVIYISLGQHFRRPPIDVEI 379
Db      252  Q---YKKKELTRPITTIQIHVSSENKD-TLLR---SKNYVIKTDIPETCDVMATNCL 303
QY      380  RRALNVHKAIO-HLLRSPDT---MVIKTENIREKYN-----DAERFSDPH----- 422
Db      304  SGINIEKCLECTLLYQNNDTSECFYVSNQVRENFNOIKAEADDEFNRNHLDTIIN 363
QY      423  -----GYOYLIIKDFDLSVSIIDAMDITIAVTNNVHPQHVAGNOINI 469
Db      464  NIKRIYKINKKEGKELLILELDFLKESITDYCKILREIDTNGP-LVNHLCNNVDV 422
QY      470  LNYI 474
Db      423  FNNLI 427

RESULT 15
S54069
Hypothetical protein YPR045c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YP9499.03c
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 05-Nov-1999
C:Accession: S54069
R:Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54059
A:Accession: S54069
A:Molecule type: DNA
A:Residues: 1-470 <BAD>
A:Cross-references: EMBL:Z49219; NID:g805025; PID:g805028; GSPDB:GN00016; MIPS:YPR045c
A:Experimental source: strain AB972
C:Genetics:
A:Gene: MIPS:YPR045c
A:Map position: 16k

```

Query Match 3.98; Score 98; DB 2; Length 470;

Best Local Similarity 19.68; Pred. No. 9.3;

Matches 80; Conservative 61; Mismatches 129; Indels 138; Gaps 17;

```

QY      86  RPTVNTTTSATSTATIL-----NPRDYCRGDQLHLLVBDHLGRKQYGG 135
Db      31  RPLDAGSFPSLITNNNNFIANOQPSLPLPRPLSMNNVSNPLAVTPLPLGLQKRINK 90
QY      146  DEL-----RARRSSPALMAGASGVTDFFNNGTYLVSTFLWEGVSLILLIHPSEGSAL 191
Db      91  NIKKKLPRVSKKASALSNGSVN-----MSNSNIYGHGAVGSASG 131
QY      192  WSAKNGYD---FVIFTGQFVNGTSQVHSECGILLTNAEL----- 229
Db      132  WKVEMGSDDELEHRRKRAERFSGPSAT-----TNSNDNLMEDEPANLANISSKSHQTD 184
QY      240  -----CY-----YLDNRDQEGFYCRP-----QHPFCALTHMYSKKNKVSYLSKQEK 272

```

```

Db      185  KRIHVGRCQTLKESYLRLLSEPNPDILRPNNILOKMYCLLMDKYOSKTAITYYLDOPK 244
QY      273  SLFERSNNGVEIEMKENTITSVSKN---TKSVDLHESGLOHOLAVDLDRNINIQOKY 329
Db      245  SM--RQDLRYOMIENSFTIKYQTHARIALENGDLQEFNOCQNR1-MALFENPTIPKSY 301
QY      330  ---CYPLIGSMITSYKMEYLTRAIDRTGGEKNTVIYISLGQHFRRPPIDVEIRRALNV 385
Db      302  SEFCY---SVLYSKLTEDY----- 318
QY      386  HKAIOHLLSPD--TMVIKTENIREKYN--NDAERFSDPHGYIQ-VL 428
Db      319  -PSISHLKLKLDGSSLELDEEHVKMIFELSDMKLVGNVHYFPKKNYL 365

```

Search completed: October 12, 2001, 16:04:29
Job time: 112 sec

Fri Oct 12 16:09:16 2001

us-09-729-454-1.rpt

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

on protein - protein search, using sw model

Run on: October 12, 2001, 16:02:37 ; Search time 34.67 Seconds

(without alignments)
850,585 Million cell updates/sec

Title: US-09-729-454-1

Perfect score: 2485

Sequence: 1 MKISMINKSLALLFLAS.....VHPQHYVGNQINILLNTIC 475

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SIDSB/gcadata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSB/gcadata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSB/gcadata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSB/gcadata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSB/gcadata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSB/gcadata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSB/gcadata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSB/gcadata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSB/gcadata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSB/gcadata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSB/gcadata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSB/gcadata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSB/gcadata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSB/gcadata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSB/gcadata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSB/gcadata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSB/gcadata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSB/gcadata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSB/gcadata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSB/gcadata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSB/gcadata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSB/gcadata/geneseq/geneseq/AA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107.5	4.3	4536	20	AAW6826
2	104.5	4.2	3923	20	AAV31237
3	104.5	4.2	4536	19	AAW41262
4	103.5	4.2	954	20	AAV33728
5	101.5	4.1	1560	10	AAW94145
6	100.5	4.0	2721	8	AAW70647
7	100	4.0	935	20	AAV31987
8	100	4.0	1031	19	AAW56102
9	99.5	4.0	705	21	AAI15553
10	99.5	4.0	705	21	AAI15562
11	99	4.0	452	21	AAV70502

12	98.5	4.0	1004	21	AAI18214	Plasmodium falciparum
13	97.5	3.9	623	20	AAV28918	Human regulatory p
14	97.5	3.9	953	20	AAV31141	Human CARD-4L prot
15	97.5	3.9	953	21	AAI15552	Apoptosis related
16	97.5	3.9	953	22	AAI20080	Human CARD-4L (lon
17	97	3.9	906	21	AAI15563	Arabidopsis thaliana
18	97	3.9	983	21	AAI15562	Arabidopsis thaliana
19	97	3.9	988	21	AAI15561	Arabidopsis thaliana
20	97	3.9	1802	21	AAI18217	Plasmodium falciparum
21	96.5	3.9	1291	19	AAW59912	Amino acid sequenc
22	92	3.7	592	17	AAI182527	Fas antigen #2.. S
23	92	3.7	1847	21	AAI182002	M. jannaschii Mj14
24	92	3.7	1847	21	AAI182002	M. jannaschii Mj14
25	91.5	3.7	648	13	AAI22551	Mouse mutant c-raf
26	91.5	3.7	1962	12	AAI10563	Mouse mutant c-raf
27	91	3.7	253	21	AAI10563	Human ORFX ORF2830
28	90.5	3.6	483	11	AAI07445	Secretory signal p
29	90.5	3.6	2325	17	AAW05590	Maize acetyl CoA c
30	90.5	3.6	2325	19	AAW05590	Maize acetyl CoA c
31	90.5	3.6	2325	19	AAW05590	Maize acetyl CoA c
32	89.5	3.6	2325	21	AAI14687	Maize acetyl CoA c
33	89.5	3.6	648	13	AAI22550	Maize acetyl CoA c
34	89.5	3.6	648	13	AAI22550	Maize acetyl CoA c
35	89.5	3.6	790	20	AAI37750	Mouse mutant c-raf
36	89.5	3.6	3418	20	AAI04355	Protein involved i
37	89	3.6	2383	21	AAI15945	Human BRCA2 (om12)
38	88.5	3.6	1521	21	AAI33736	E. coli proliferat
39	88.5	3.6	1521	21	AAI33736	Agarase 4-3 protei
40	88.5	3.6	1521	21	AAI33736	Arabidopsis thaliana
41	88.5	3.6	1521	21	AAI33736	Arabidopsis thaliana
42	88.5	3.6	1703	21	AAI33736	Arabidopsis thaliana
43	88.5	3.6	1740	21	AAI33736	Arabidopsis thaliana
44	87.5	3.5	648	13	AAI22553	Arabidopsis thaliana
45	87.5	3.5	648	20	AAI0668	Mouse mutant c-raf

ALIGNMENTS

RESULT 1	AAW6826	standard; protein; 4536 AA.
ID	AAW6826	
XX	AAW6826	
AC	AAW6826	
DT	22-APR-1999	(first entry)
XX		
DE	Amino acid sequence of human apolipoprotein B-100 (apob-100).	
XX		
KW	Human apolipoprotein B-100; very-low density lipoprotein; VLDL;	
KW	apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;	
KW	nuclear localization sequence; gene therapy; cancer; cystic fibrosis;	
KW	non-small cell lung carcinoma; diabetes; arteriosclerosis.	
XX		
OS	Homo sapiens.	
XX		
PN	W09656938	AI.
XX		
PD	17-DEC-1998.	
XX		
PF	10-JUN-1998;	98MO-US11927.
XX		
PR	14-MAY-1998;	98US-0079030.
PR	13-JUN-1997;	97US-0874807.
XX		
PA	(BAYU) BAYLOR COLLEGE MEDICINE.	
XX		
PI	Guevara JG,	Hoogveen RC, Moore JP;
XX		
PT	WPI, 1999-070331/06.	
PT	Composition comprising nucleic acid bound to LDL or VLDL lipoprotein	
PT	- used for delivering nucleic acid to cells for gene therapy and	

PT antisense treatment
 XX
 PS Disclosure; Fig 1A-C; 293pp; English.
 XX
 CC The present sequence represents human apolipoprotein B-100 (apoB-100).
 CC apoB-100 is a major apoprotein component of very-low density
 CC lipoproteins (VLDL). Intermediate density lipoprotein (IDL), low
 CC density lipoproteins (LDL) and lipoprotein a. The specification
 CC describes a composition that comprises LDL and apolipoproteins for the
 CC binding and in vivo transport of nucleic acids. Binding domains
 CC (see AAW96878-97) and nuclear localization sequences (see AAW96878-97)
 CC for use in the composition can be derived from the present sequence. The
 CC composition is used to deliver nucleic acids to eukaryotic cells, in
 CC vivo or in vitro, for expressing a therapeutic polypeptide or antisense
 CC molecule (or ribozyme). Specifically they are used for gene therapy of
 CC cancers (particularly non-small cell lung carcinoma), diabetes, cystic
 CC fibrosis and arteriosclerosis.
 XX
 SQ Sequence 4536 AA;

Query Match 4.3%; Score 107.5; DB 20; Length 4536;
 Best Local Similarity 20.7%; Pred. No. 5.1;
 Matches 90; Conservative 71; Mismatches 169; Indels 105; Gaps 22;

QY 56 KPLISLKPLTETE-LRIKETIEKLDQOIPPPPTHV---NTTSATHS-TATILNPRD 109
 DB 1998 kyp11sepin1ida1emrdaevk-----pqet1vafvkykngdvhs1n1pfet1q 2051
 QY 110 TYCRGDQHLILEVRDHLGRKOYGD-FLRARKSSPALMAGSKGVDFNNGTYIVSFT 168
 DB 2052 eyfernrqclivlenvgvnrlnkhindqfvk-----yralajklpqqan-dylnsfn 2103
 QY 169 LFWGVGSLSLLIHPSGVSAALNS-----ARNQYDRAVF 204
 DB 2104 --wergvs-----hakekitaltkkyt1lendq1aladakinfnek1sq1q1y--m1q 2153
 QY 205 TCGFVNGTSOVHSECGILNTNAELCOYLNRDQEGFVCVRPQHPMPCALTHMYSK---- 260
 DB 2154 fdqykdsydlhd1k1a1n1id1e1ek1s1deh--yhlrvn1vkt1hd1h1f1e1nd1 2211
 QY 261 NKKVSYLSKQKSLFERSNNGVEIMEKNNTSVSKNTLKSVDL-HESGKL-OHQLAVDL 318
 DB 2212 nksqstasw1qnvdtkyq1r1q1ek1q1k1-----rh1qnd1q1h1a1q1k1h1e1ad1v 2267
 QY 319 DRNINIQOKTCYPLGSMNTSVKEMEYLRATDRTGKKNVTIVISLGOHFPPPT--- 375
 DB 2268 -----tv11dq1q1t-----cl1sfer1----ndv1-----ehvkh1v1n1l1 2298
 QY 376 -DVFTRALNVHKAIOHLLRSPT-----MVIKTENIREMYNDARSPDRHGVIQVLI 429
 DB 2299 gdfveaek1naf1akv1h1e1ryevdq1q1v1mdk1v1e1t1h1q1k1ek1s1n1v1q1v1k 2358
 QY 430 IKDIFQDLSVSIIDA 444
 DB 2359 lkydfek1vg1f1dda 2373

RESULT 2
 AAY31237
 ID AAY31237 standard; Protein; 3923 AA.
 XX
 AC AAY31237;
 XX
 DT 08-NOV-1999 (first entry)
 XX
 DE Human Apo B protein fragment.
 XX
 KW Apo B; hyperlipidemia; human; treatment; hepatocyte; apoprotein;
 KW Apo A1; low density lipoprotein; LDL; blood; therapy; arteriosclerosis;
 KW high density lipoprotein; HDL; cholesterol; coronary heart disease;
 KW Alzheimer's disease; hypobetalipoproteinemia; dysbetalipoproteinemia.
 XX

OS Homo sapiens.
 FH Key Location/Qualifiers
 FT 1..3923 Protein
 FT /label= Apo B
 FT /note= "Partial sequence, no start codon given"
 PN MO9940789-A1.
 PD 19-AUG-1999.
 PE 28-AUG-1998; 98MO-US17908.
 PF 30-JUN-1998; 98US-0108006.
 PR 12-FEB-1998; 98US-0074497.
 PA (MINI) UNIV MINNESOTA.
 PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 PI Bandyopadhyay PT, Kren BT, Roy-Chowdhury J, Steer CJ.
 DR WPI: 1999-527333/44.
 DR N-PSDB: AA209525.
 XX
 PT Mutating apolipoprotein genes in hepatocytes to control cholesterol
 PT levels, e.g. for treating or preventing hyperlipidemia, particularly
 PT arteriosclerosis
 PS Claim 4: Page 75-83; 106pp; English.
 XX
 CC This invention describes a novel method for the genetic treatment of
 CC hyperlipidemia by altering genes in hepatocytes, for apoprotein (apo) B,
 CC E or A1. Low density lipoprotein (LDL) levels in the blood are reduced by
 CC altering an apo B gene (I) in a hepatocyte. The invention describes a
 CC method for the therapeutic and/or prophylactic method involving altering
 CC an apo E gene in hepatocytes by introducing the mutations Arg12Cys,
 CC Arg158Cys or Cys158Arg and a method for ameliorating arteriosclerosis by
 CC altering the apo A1 gene in a hepatocyte so that the altered protein can
 CC downregulate. Altering expression of apo genes regulates levels of high and
 CC low density lipoprotein cholesterol. Altering expression of apo B, E and
 CC A1 genes is used to treat or prevent arteriosclerosis, coronary heart
 CC disease, Alzheimer's disease, hypobetalipoproteinemia, and
 CC dysbetalipoproteinemia. This sequence represents a fragment of the human
 CC Apo B protein described in the method of the invention.
 XX
 SQ Sequence 3923 AA;

Query Match 4.2%; Score 104.5; DB 20; Length 3923;
 Best Local Similarity 20.5%; Pred. No. 7.8;
 Matches 89; Conservative 72; Mismatches 169; Indels 105; Gaps 22;

QY 56 KPLISLKPLTETE-LRIKETIEKLDQOIPPPPTHV---NTTSATHS-TATILNPRD 109
 DB 1385 kyp11sepin1ida1emrdaevk-----pqet1vafvkykngdvhs1n1pfet1q 1438
 QY 110 TYCRGDQHLILEVRDHLGRKOYGD-FLRARKSSPALMAGSKGVDFNNGTYIVSFT 168
 DB 1439 eyfernrqclivlenvgvnrlnkhindqfvk-----yralajklpqqan-dylnsfn 1490
 QY 169 LFWGVGSLSLLIHPSGVSAALNS-----ARNQYDRAVF 204
 DB 1491 --wergvs-----hakekitaltkkyt1lendq1aladakinfnek1sq1q1y--m1q 1540
 QY 205 TCGFVNGTSOVHSECGILNTNAELCOYLNRDQEGFVCVRPQHPMPCALTHMYSK---- 260
 DB 1541 fdqykdsydlhd1k1a1n1id1e1ek1s1deh--yhlrvn1vkt1hd1h1f1e1nd1 1598
 QY 261 NKKVSYLSKQKSLFERSNNGVEIMEKNNTSVSKNTLKSVDL-HESGKL-OHQLAVDL 318
 DB 1599 nksqstasw1qnvdtkyq1r1q1ek1q1k1-----rh1qnd1q1h1a1q1k1h1e1ad1v 1654
 QY 319 DRNINIQOKTCYPLGSMNTSVKEMEYLRATDRTGKKNVTIVISLGOHFPPPT--- 375

CC This sequence represents a 954 amino acid insecticidal
 CC toxin from *Photobabidus luminescens*. It is one of three insecticidal
 CC toxins (AA93728-Y33730) encoded by open reading frames (ORFs)
 CC in a 38kb fragment of *P. luminescens* DNA (AA206851). This sequence
 CC is encoded by ORF5. *P. luminescens* is a member of the
 CC Enterobacteriaceae family and is a symbiotic bacterium of
 CC nematodes of the genus *Heterorhabdus*. The nematodes colonize insect
 CC larvae, kill them, and their offspring feed on the dead larvae. However,
 CC the insecticidal agents are produced by *P. luminescens* rather than the
 CC nematodes. The toxins have activity against Lepidopteran insects such as
 CC Cabbage Looper (*Trichoplusia ni*), European Corn Borer (*Ostrinia
 CC nubilalis*) and Fall Armyworm (*Spodoptera frugiperda*) and also against
 CC Coleopteran insects (e.g., Colorado Potato Beetle, *Leptinotarsa
 CC decemlineata*). In addition the toxins are active against strains
 CC resistant to known insecticides. The DNA sequence can be used to generate
 CC transgenic plants of various species that are resistant to economically
 CC important insect pests and also for recombinant production of the toxins
 CC for use as insecticides.

XX Sequence 954 AA:

Query Match 4.28; Score 103.5; DB 20; Length 954;
 Best Local Similarity 20.18; Pred. No. 0.93;
 Matches 96; Conservative 80; Mismatches 158; Indels 143; Gaps 24;

QY 33 VMSALNLSISLHMNNSTKSLFPTPLI--SLKPLTETELRIKEIEKIDQIPRPFTH 90
 DB 325 VLSIRNDAESTRFH--qkwapenlytdslyqlslstgemanlqgsgndlp----- 376
 QY 91 VNTTSTATSTATILNPRDYCRGDQHLHLEVRDHLGRKQYCGPLFARNSSPALMAG 150
 DB 377 -sltpedmtyenytytydyr-----ggnltkqhspsatqnn 415
 QY 151 ASGVYTDENNGTYLVSTLFWEGVSLSLIHSEGVSAIWSARQGYDRVFTGQFVN 210
 DB 416 ytlitlvansrnray-----lsltleddpaq-vdalida--ggnqntllsgnln 461
 QY 211 GTSQVSECGLLNFMNLEOXYD--NRQGEF-----YCARPQHMCAA 253
 DB 462 -----wnrtgel-qhncvlvkrckganddreyryssdgrllklineqg----- 503
 QY 254 LTHWYSKNNKVSYSLSKQKSLFERSNVGVEIEMKENTISVSKNTLKSVDLH--ESGKLQ- 311
 DB 504 -tsasqgtqrltlypslelrltqstltedqg--vltvgasgragrvlhwdsqged 559
 QY 312 ---HOL-----AYLDENINIQOKQCYPLIGSMITYSV--KEMEYLTRAIDRTGG 356
 DB 560 ldmqqltyaydnllgsqgltdskgeislseeyypygatawatkrteasykcltysgk 619
 QY 357 EKNTVIVISLG-QHFRFP-----PIDVFIIRALNVHKAIOH--LLRSPDTMVI 402
 DB 620 erdatgluyyryyqpwgvrwlsadpaqtdv-----glnlyymvrnnpvllldpdl 672
 QY 403 IKTENIREMYNDAREFSDPHGYQYIILKIDIFODLSYIIDADITIAVGNNVHP 459
 DB 673 -----mptlaetia-----alqknkvadapsrlnatval-----nltrp 708

RESULT 5

AA94145 AAP94145 standard; protein; 1560 AA.

XX AC AAP94145;
 XX DT 07-JUN-1990 (first entry)
 XX DE S. cremoris protease.
 XX KM Lactic acid bacteria; cheese; *Streptococcus cremoris* SK112;
 XX KN protease; pSK112; chymosin; prochymosin.
 XX PN NL8701378-A.

XX PD 02-JAN-1989.
 XX PF 12-JUN-1987; 87NL-0001378.
 XX PR 12-JUN-1987; 87NL-0001378.
 XX PA (NEZU-) NEDERL INS ZUIVELON.
 XX PT Simons AFM, De Vos WM;
 XX DR WPI: 1989-030097/04.
 XX NR-PSDB: AAN91159.
 PT DNA fragment having region specific for lactic acid bacteria
 PT is contained in plasmid in microorganism used in prodn. of
 PT protein and food prodn. eg cheese.
 XX Claim 6; fig 7a; 43pp; Dutch.

XX Sequence 1560 AA:

Query Match 4.18; Score 101.5; LB 10; Length 1560;
 Best Local Similarity 21.18; Pred. No. 3.3; Indels 131; Gaps 21;
 Matches 96; Conservative 52; Mismatches 177;

QY 58 PLISLKPLETEBLRIK-----EIEKIDQIPRPFTHVNTTSTATILNPK 108
 DB 714 paveldkfstidkfkilftnrltelygnds-----ndtaaytsatdpn-- 761
 QY 109 DTYCRGDQHLHLEVRDHLGRKQYCGPLFARNSSPALMASGVYTDENNGTYLVST 168
 DB 762 -----sgvlydkkldgaalaaqsn--ltvpnqkkaaleit 794
 QY 169 LFWEGVSLSLIHSEGVSAIWSARQGYDRVFTGQFVNCTSVV--HSECG----- 220
 DB 795 I-----slpkslnlpymgffgqdm-----dqkivdslnqtlvyspaqundtvp 837
 QY 221 LIINTVAELCOYLDN--ROQEGYCVPRQHMPCALTHMNSKNNKVSYSLSKQKSLFERS 278
 DB 838 llnkntg-tlyygmwtddagngvddq-----aaisdkua; yndlsnkylitrus 891
 QY 279 NQGVETME-----KNTISVSKNTLKSVDLHESKRLQBOIAV-----DLDEMINIOW 927
 DB 892 nqvvdlldgqknkvtlsselnkkylnaasqgylynapawdqttydgrdnktaad 951
 QY 328 KICYPLIGSMITYSVKEMEYLTRAIDRTGEEKNTVIVISLGQHPFPIDVFI-----RR 981
 DB 952 -----gslytyrisgyp-----eggdgrkfytdv-----piklskqpw-40kky 999
 QY 382 ALNVHKAIOHLLRSPDTKWITIKTENIR-ENYNDAREFSLPHGYQYIILKIDFODL 437
 DB 990 vtdvpfkldskaprvpvalssaktengtktyltaeakdlsqatdkvktvtoleevtl 1049
 QY 438 SVSIIDADITIAVGNNVHP-----OHVVGNOIN 468
 DB 1050 datftdag--ltadgykltetrlpsdeqagailngdn 1083

RESULT 6

AA970647 AAP70647 standard; protein; 2721 AA.

```

AC  AAP70647:
XX
XX  17-APR-1991 (first entry)
XX
XX  Sequence of N-terminal apolipoprotein B (apoB).
XX
XX  Lipid-binding peptide; apolipoprotein purification.
XX
XX  Homo sapiens.
XX
XX  Key      Location/Qualifiers
XX  Feature  87..167
XX  FT       /label=Signal
XX  FT       Protein      168..8247
XX
XX  W08702062-A.
XX
XX  09-APR-1987.
XX
XX  02-OCT-1986;      86WO-0502075.
XX
XX  04-DEC-1985;      85US-0804692.
XX  04-OCT-1985;      85US-0784418.
XX  27-FEB-1986;      86US-0834300.
XX
XX  (BIOT-) BIOTECHN RES PARTN.
XX
XX  Protter AA, Vigne JL, Mallory JB, Talmadge KD;
XX  WP1: 1987-108703/15.
XX  DR  N-PSDB; AAN70996.
XX
XX  Purified lipid-binding peptide prodn. - by expression of
XX  recombinant system and adding lipid to chcd. peptide mixt. to
XX  form lipo-peptide complex
XX
XX  Claim 50; pp73-85; 123pp; English.
XX
XX  ApoB is a lipid-binding peptide adapted to bind to phospholipids at
XX  one or more amphipathic alpha-helical peptide regions. ApoB gene is
XX  pref. expressed in Chinese hamster ovary (CHO) cells and the gene is
XX  introduced in an expression vector having a regulatable promoter
XX  derived from the human metallothionein II gene.
XX
XX  Sequence: 2721 AA:

Query Match      4.0%; Score 100.5; DB 8; Length 2721;
Best Local Similarity 20.5%; Pred. No. 10;
Matches 89; Conservative 71; Mismatches 170; Indels 105; Gaps 22;

YY  56 KTRPLSLKPLTEFE-LRIKEIIEKLDQIPRPPTHV-----NTTSATHS-TATILNPD 109
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  4025 krpililsepililidaletmdavek-----pqeltlvaivkydkqdvshnlpffeilq 2078

YY  110 TYCGKDOLHILLEVRDHLGRROYGD-FLRRKSPALMAGSGVTDENNGTYLVST 168
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  2079 eyfernqrqtlivlenvqnllklnldqfvrk-----yralaklpqgan-dynlsfn 2130

YY  169 LFMEGVSLSLLIHPSSEGVSLMS-----ARNGYDRVIF 204
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  2131 --weryvs-----hakekitaltkkrytlendiqaladakinfeklsglqly--mliq 2180

YY  405 TQDFVNGTSGVHSEGLILNTNAELCOYLDRQDEGFCVPRQHPICALITHMSK---- 260
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  2161 idqylkdsydfhqlklalaniaidelfeklsidsh--yhlrvnlvxxhldhlfienidf 2238

YY  261 NKKVSYLSKQEKSLFPRSNVGVIEKEFNITSXKNTLAKVDL-HESGCL-OHQLAVDL 318
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  2259 nksqsstasawlqnvdykkyqitlqikeklqyk-----rhinqldlqnhlaqklkghlaadlv 2294

YY  319 DRNINIMOKYCYPLIGSNYSYVAKEMEXLTRAIDRTGGEKNTVIVISLGOHPFPPI--- 375
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

DB  2295 -----rvllldqigt-----tlseferl-----ndvl-----ehvkhfvlnpy 2325

YY  376 -DVFIRRALNVHKAIOHLHLLRSPDT-----MWIKTENIRPMYNDAREFSDPHGYIYLI 429
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  2326 wdfewaeeklnafraqkbelierveyedqdhqvlmcklvelanqyklketqlksnvlqyvk 2385

YY  430 IKDIFODLSVSIIDA 444
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  2386 lkdyfeklvvgfidda 2400

RESULT 7
AA31987
ID  AA31987 standard; Protein; 935 AA.
XX
XX  AA31987;
XX
XX  21-DEC-1999 (first entry)
XX
XX  DE  Alpha-ketoglutarate dehydrogenase E1 subunit.
XX
XX  KW  Glutamic acid; alpha-ketoglutarate dehydrogenase; AKGDH; sucAB;
XX  sucA.
XX
XX  OS  Enterobacter agglomerans.
XX
XX  PN  EP952221-AA2.
XX
XX  PD  27-OCT-1999.
XX
XX  PF  17-MAR-1999; 99EP-0105508.
XX
XX  PR  18-MAR-1998; 98JP-0069068.
XX  19-OCT-1998; 98JP-0297129.
XX
XX  PA  (AJIN ) AJINOMOTO KK.
XX
XX  PI  Izut H, Ono E, Matsui K, Moriya M, Ito H, Hara Y;
XX  WP1: 1999-582803/50.
XX  DR  N-PSDB; AA20083.
XX
XX  Novel bacteria used for production of L-glutamic acid -
XX
XX  Example 3; Page 19-22; 32pp; English.
XX
XX  The present sequence represents the Enterobacter agglomerans
XX  AJ13355 (FERM P-16644) alpha-ketoglutarate dehydrogenase (AKGDH)
XX  E1 subunit. The sequence was deduced from the sucA gene
XX  identified in a cloned fragment (see AA220083) of AJ13355 DNA.
XX  This cloned fragment also included the such open reading frame
XX  encoding the AKGDH E2 subunit (see AA31987). The invention provides
XX  Enterobacter and Serratia microorganisms that have an improved
XX  ability to produce L-glutamic acid as a result of an increase in
XX  activity of an enzyme catalysing a reaction for L-glutamic acid
XX  biosynthesis and/or a decrease or deficiency in an activity of an
XX  enzyme that catalyses a reaction branching away from a pathway for
XX  L-glutamic acid biosynthesis, such as AKGDH. An E. agglomerans
XX  strain deficient in AKGDH was produced using the sucAB gene by
XX  homologous recombination.
XX
XX  Sequence 935 AA:

Query Match      4.0%; Score 100; DB 20; Length 935;
Best Local Similarity 20.3%; Pred. No. 2;
Matches 86; Conservative 42; Mismatches 136; Indels 160; Gaps 23;

YY  34 MSA-LNTLSISLHYWNSTSLFPPKPLISLKPLTELELRKEIIEKLDQIPRPPTHVN 92
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB  535 wpyjnh-----hewde-----pypaqvdmkrlkelalrlsqvpeqlvq----- 573

YY  93 TTTSATHTSTATILNPDRTYCRQDOLHILLEVRDHLGRROYGD--DFLRRKSPALMAG 150

```


Db 386 KKKKyeLlnkHbllnllkIntrelswmyeLsahfyyfIdheni-----yVlWkll 440
 QY 431 KQIFQDLSTSTI 442
 Db 441 twltdlvsall 452

RESULT 9
 AAB15559
 ID AAB15559 standard; Protein: 705 AA.
 XX

AC AAB15559;
 DT 28-FEB-2001 (first entry)
 XX

DE Apoptosis related protein encoded by gene 2 clone HDPB68.
 XX

KM Cytostatic; antiinflammatory; immunosuppressive; antisclerotic; cardiant;
 KM virucidal; anti-AIDS; vasotropic; anti-Ischaemic; antiParkinsonian;
 KM anti-Alzheimer; gene therapy; human; apoptosis; fusion protein; cancer;
 KM colon; breast; prostate; melanoma; lymphoma; inflammation; herpes;
 KM autoimmune disorder; multiple sclerosis; viral infection.
 XX

OS Homo sapiens.
 XX

PN W0200056752-A2.
 XX

PD 28-SEP-2000.
 XX

TF 15-MAR-2000; 2000WO-US06642.
 XX

FR 24-MAR-1999; 99US-0126018.
 XX

PR 17-JUN-1999; 99US-0139638.
 XX

PR 18-AUG-1999; 99US-0149449.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX

FI Ruben SM, NI J, Young PA;
 XX

DK WPI: 2000-587660/55.
 XX

DK N-PSDB: AAB155798
 XX

PT Nucleic acids encoding human apoptosis associated protein, useful for
 the prevention, treatment and diagnosis of e.g. Alzheimer's and
 Parkinson's disease, inflammation and ischaemic injury -
 XX

PS Claim 11: Page 265-267; 273pp; English.
 XX

CC The invention relates to the isolation of genes encoding 9 human
 CC apoptosis-related proteins. The nucleotide sequences AAB155790-A95798
 CC encode the human apoptosis related proteins AAB15551-B15559. The genes
 CC can be used to generate fusion proteins by linking to the gene for the
 CC human immunoglobulin G Fc (196 Fc) portion (AAB155799) for increasing the
 CC stability of the fusion protein as compared to the human protein only.
 CC The gene and encoded protein may be used in the prevention, treatment
 CC and diagnosis of diseases associated with inappropriate apoptosis
 CC associated protein expression, e.g. cancer (e.g. colon, breast and
 CC prostate cancer, melanomas and lymphomas), inflammation, autoimmune
 CC disorders (e.g. multiple sclerosis) and viral infections (e.g. herpes)).
 XX

SU Sequence 705 AA:
 XX

Query Match 4.0%; Score 99.5; DB 21; Length 705;
 Best Local Similarity 20.5%; Pred. No. 14;
 Matches 96; Conservative 58; Mismatches 154; Indels 161; Gaps 23;

QY 18 IASWITFVFNSTKVMASLNLSTLHYMNWNSKSLFPTPLISAKPLTELRKEITIE 77
 Db 144 IICWILITFCF-----H-----RADEGSPQIPQCTMLTLDVFI--VLE 182
 QY 78 KLDQOIPRPPTHVNTTTSATHTATILNPDITYC-----RGDQLHILLEVRDLG-- 128

Db 183 vhlntmpssalvgtrtrs-----pyeclhagridtLcslsqvahnqsmekajlftvtgeevxas 238
 QY 129 -----RRKOYGGDPLRRMRSSPALMAAGSGVTDPNKNGYLVSTLFW---EGQVSLILL 181
 Db 239 glqerdmqlg--flfa---lpeqlpgqdgqxyefhlqlgaftaflvlddrvtgqell 293
 QY 182 IHPSEGVSAIWSARNGY-----DRVFTGGFVNGTSQVHSE 218
 Db 294 rffgevmppagaattscypflfpfcqlqsgparedlfxkndhfqtnlfl----- 344
 QY 219 GGLLTNMAELCOYLIDNRDQGFVCPVPHMCAAL-----THMSKMKVSYLSKO 270
 Db 345 cgliskakqkllfhl-----vpaalrlrkxkalwalfsfl--gyllsl 387
 QY 271 EKSLFERSNGVEIMEKF-----NTISVSKCNTLK----- 300
 Db 388 prvgvesfn-qvampftlmlrclyetqsgkvqglaaqgicanyklkycnacsadcsa 446
 QY 301 -SYDLHESGKLQHOLAVDLRN-----INIQOKYCYPLIGSMITYSKENEYLTTRALDRTG 355
 Db 447 lsfvlhfpk-----rlaldnnlnndygvrelqpctsrflvrlsvnql-----tdg 495
 QY 356 GER-----NTVIYISLQGHFRFPPIDV---FIRRALNWKAIQHILL 394
 Db 496 gkvvlseeltkykklvtlyglgnql-tdvgarvylklldeckgltlhlk 543

RESULT 10
 AAB15562
 ID AAB15562 standard; peptide; 705 AA.
 XX

AC AAB15562;
 XX

DT 28-FEB-2001 (first entry)
 XX

DE Fragment of apoptosis related protein encoded by gene 2 clone HDPB68.
 XX

KM Cytostatic; antiinflammatory; immunosuppressive; antisclerotic; cardiant;
 KM virucidal; anti-AIDS; vasotropic; anti-Ischaemic; antiParkinsonian;
 KM anti-Alzheimer; gene therapy; human; apoptosis; fusion protein; cancer;
 KM colon; breast; prostate; melanoma; lymphoma; inflammation; herpes;
 KM autoimmune disorder; multiple sclerosis; viral infection.
 XX

OS Homo sapiens.
 XX

PN W0200056752-A2.
 XX

PD 28-SEP-2000.
 XX

TF 15-MAR-2000; 2000WO-US06642.
 XX

FR 24-MAR-1999; 99US-0126018.
 XX

PR 17-JUN-1999; 99US-0139638.
 XX

PR 18-AUG-1999; 99US-0149449.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX

FI Ruben SM, NI J, Young PA;
 XX

DK WPI: 2000-587660/55.
 XX

DK N-PSDB: AAB155798
 XX

PT Nucleic acids encoding human apoptosis associated protein, useful for
 the prevention, treatment and diagnosis of e.g. Alzheimer's and
 Parkinson's disease, inflammation and ischaemic injury -
 XX

PS Disclosure: Page 20-21; 273pp; English.
 XX

CC The invention relates to the isolation of genes encoding 9 human
 CC apoptosis-related proteins. The nucleotide sequences AAB155790-A95798
 CC encode the human apoptosis related proteins AAB15551-B15559. This
 CC sequence represents a fragment of the protein encoded by the gene given
 in the descriptor line. The genes can be used to generate fusion

CC osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis,
 CC scleroderma, Sjogren's syndrome, and autoimmune thyroiditis,
 CC complications of cancer, hemodialysis, extracorporeal circulation; viral,
 CC bacterial, fungal, parasitic, protozoan, and helminthic infections; and
 CC trauma. The HRP polynucleotide may be used for diagnosis of these
 CC conditions, and as a source of primers and probes.

XX
 SQ Sequence 623 AA:

Query Match 3.9%; Score 97.5; DB 20; Length 623;
 Best Local Similarity 19.4%; Pred. No. 1.8; Mismatches 172; Indels 229; Gaps 34;
 Matches 116; Conservative 82;

```

13 ALLEFIASWTFYFONSTKVMASALNLSLHWNSTKSLPPTPL-----59
DB 72 avcglas-----hpnstdv-hlnlnstlftf-----gqellsdcklelnsgrryllgl 119
OY 60 -----ISLPLTERELRIKEITE-KLDQOIPRPPT-----HYNTTTSATHTSTATIL 105
DB 120 nglgkmlslsagkrvprphndlyhlrtrempsdktphcymevdteramlekeaeel 179
OY 106 NPPDTGCGDDQLHILEVRDHLGRKQGGDFLRARNSPALMAGAS-----GKVTDF 158
DB 180 ahedaec-eklmelyerleel-----dadkaemraarllhglgftpamgkklkldf 229
OY 159 NNGTYLVSFTLEWEGVSL-----LLLIHPSGV---SALMSARN-OGYDRI-- 203
DB 230 sgg-----wrmvalaralfirpmlldpctnhldacvyleeektkrlivl 280
OY 204 --FTGQFVNG--TSQVHSECGILNTNMLECOYLDNRDQ-----EGF----- 241
DB 281 vshsqdflngvctnlhmh-----nklklygnydqyvktrtleengmkrfhwgd 333
OY 242 -----YCVRPQH-----MPCALTFHMSKKNKVSY----- 266
DB 334 glahmknylarfghgsaklarqgsketklqmasgltetrvsdktlsfyfpcgklpp 393
OY 267 -----LSKQESLPEKSNVGVIMKFNIT---SVSKCMTLK-----S 301
DB 394 pvlmwgnvsfkytkdpcplynlefqldtrvalvpgnagksllklitgelllptdm 453
OY 302 VDLHESGK-----LQHQLAVDLDRNINIQMOKYCPPLIGSNTYSVKEMYLTRAIDRT 354
DB 454 lfrhshvklgrfghlqeqdlldls-pleymmkcp-----elkekeemrkllgry 503
OY 355 G--GEKNIVY--ISLQGNR-----PPIDVIFIRALNVKAIQHLHLSPTDWT 402
DB 504 gltgqgvsplrnlsdqkrcvclawlawqpnlllfdpfn-----hldletdal-- 555
OY 403 IKTENIREMYNDAREFSDFHGYT-----QYLIIKQIFODLSY---SIIDAW-DITIAV 451
DB 556 -----adaInetegmmllvshdflllqyagelwcekgqtltkwpgdl-1ay 601

```

RESULT 14

AAV31141
 ID AAV31141 standard; Protein; 953 AA.

AC AAV31141;

DT 25-OCT-1999 (first entry)

DE Human CARD-4L protein.

XX CARD-3; caspase recruitment domain; CARD-4; regulation; detection;
 XX caspase activation; detection; screening; therapy; diagnosis; disease;
 XX apoptotic cell death; Fas/APO-1 receptor complex; TNF receptor complex;
 XX cancer; follicular lymphoma; carcinoma; p53 mutation; viral infection;
 XX hormone-dependent tumour; autoimmune disorder; Alzheimer's disease;
 XX systemic lupus erythematosus; immune-mediated glomerulonephritis; stroke;
 XX Parkinson's disease; amyotrophic lateral sclerosis; retinitis pigmentosa;
 XX spinal muscular dystrophy; cerebellar degeneration; anaemia; drug;

KM myelodysplastic syndrome; myocardial infarction; cell proliferation;
 KM cell differentiation; cell survival; CARD-4L; CARD-4S; CARD-4Y;
 KM CARD-4Z; human.

XX Homo sapiens.

XX Key Location/Qualifiers

```

FH 15..114 /note="predicted CARD domain"
FH Domain /note="predicted kinase 1a (P-loop) domain"
FH 197..212 /note="predicted kinase 1a (P-loop) domain"
FH 198..397 /note="nucleotide binding domain"
FH Region /note="nucleotide binding domain"
FH 202..209 /note="predicted Walker Box A"
FH 273..288 /note="predicted kinase 2 domain"
FH 280..284 /note="predicted kinase 2 domain"
FH 327..338 /note="predicted Walker Box B"
FH Domain /note="predicted kinase 3a subdomain"
FH 674..950 /note="region containing predicted leucine-rich"
FH 674..701 /note="predicted leucine-rich repeat 1"
FH 702..727 /note="predicted leucine-rich repeat 2"
FH 728..754 /note="predicted leucine-rich repeat 3"
FH 755..782 /note="predicted leucine-rich repeat 4"
FH 783..810 /note="predicted leucine-rich repeat 5"
FH 811..838 /note="predicted leucine-rich repeat 6"
FH 839..866 /note="predicted leucine-rich repeat 7"
FH 867..894 /note="predicted leucine-rich repeat 8"
FH 895..922 /note="predicted leucine-rich repeat 9"
FH 923..950 /note="predicted leucine-rich repeat 10"

```

XX MO9940102-A1.

XX 12-AUG-1999.

XX 05-FEB-1999; 99MO-US02544.

XX 08-DEC-1998; 98US-0207359.

XX 06-FEB-1998; 98US-0019942.

XX 17-JUN-1998; 98US-0099041.

XX (MILL-) MILLENNIUM PHARM INC.

XX Bertin J;

XX MPI: 1999-494269/41.

XX N-PSDB: AA209247.

XX Novel CARD-3 and CARD-4 genes and polypeptides used or treating

XX survival

XX Claim 13a; Fig 4; 18pp; English.

XX This invention describes the isolation of novel human caspase
 XX recruitment domain, CARD-3 and CARD-4 polynucleotides and proteins of
 XX partial murine CARD-4L protein and genes. The genes and proteins of
 XX the invention are involved in the regulation of caspase activation.
 XX The caspase recruitment domain (CARD) polynucleotides, polypeptides,
 XX homologues and antibodies can be used in screening assays, detection

assays, predictive medicine and therapeutic and prophylactic methods of treatment. The methods may be used to diagnose and treat patients which are suffering from a disorder associated with abnormal level or rate of apoptotic cell death, abnormal activity of the Fas/Abp-1 receptor complex, abnormal activity of the TNF receptor complex, or abnormal activity of a caspase. Diseases that may be treated include cancer (particularly follicular lymphoma, carcinomas associated with mutations in p53 and hormone-dependent tumours), autoimmune disorders (e.g. systemic lupus erythematosus, immune-mediated glomerulonephritis), viral infections, Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, retinitis pigmentosa, spinal muscular dystrophy, cerebellar degeneration, anaemia, myelodysplastic syndrome, myocardial infarction, and stroke. CARD-1 protein interacts with other cellular proteins, and so can be used for regulation of cellular proliferation and differentiation (and cell survival). The CARD proteins may also be used to for screen drugs or compounds which modulate their activity. The CARD-4 gene can express a long transcript that encodes CARD-4L, a short transcript that encodes CARD-4S or two CARD-4 splice variants, CARD-4Y and CARD-4Z. This sequence represents the human CARD-4L protein described in the method of the invention.

Sequence 953 AA:

Query Match 3.98; Score 97.5; DB 20; Length 953;
Best Local Similarity 20.58; Pred. No. 3.6;
Matches 96; Conservative 58; Mismatches 154; Indels 161; Gaps 23;

```

0Y 16 LKSWIEFVFNQSNKRWASALNLSLHWNWNSYSEFPPTLSLKLTERELRIKITE 77
Db 492 IICWIIFFICFQ-----HL-----raefgsppqjpcctctldvll-vte 433
0Y 78 KLDQOIPREFHVNNTTSATNSTATILNPRDYC-----RGDQHLILEVDHL-- 127
Db 431 vlnhmrgssivgnrts-----pvelthagdtlcsigvahrimgxelslftfgeeqas 48
0Y 128 ---GRRKUYGSDFLRRARSSPALMGASGKTFDFNNGTYLVSEFLFW---EGVYSLL 18
Db 487 glqetdmglg--flra--lpeIsp9qdgqyefthltaqafitflviddrvgtejll 54
0Y 182 IHPSEGVASLMSARNQY-----GRVFTGQFNGHSGVHSE 21
Db 542 rffgwpmpqqaatiscyrrflpfcqlqgspparedfknkhgffthllf----- 59
0Y 219 GCLLINTNAELCOYLDNRQEGFYCVRPDHPCCAL-----THMSKNKRYSLSKQ 27
Db 593 gyltskaxqkllhll-----vpaalrrkklwaihllsslr--gylkal 63
0Y 271 EKSLPERSNGVELIMEKF-----NHSKCNILK----- 30
Db 636 prvyqvyrsh-qvqamptfkmrlrcteyetqskvqglaaqglcaanykltycncasdcsa 69
0Y 301 -SVDLHESGKLQHLAVDLDRN---INIQOMQYCYPLIGSMYVSYKMEYTRAIADRG 351
Db 695 lsfvhlhprk--rlaldldmnnldygvrelqpcsrllvrlslsnql-----ctg 74
0Y 356 GER-----NTVYVSLCOGHRPRPIDV---FIRALVNHKAIOHLL 394
Db 744 gkxkylseelckykivlyajlymqrl--tdvayarvylldcecgqtlhkl 791

```

RESULT 15
AAB15552
10 AAB15552 standard; Protein; 953 AA

A0: AAB15552;

DT 28-FEB-2001 (first entry)

Apoptosis related protein encoded by gene 2 clone HDPBW68

KM Cytostatic; antineoplastic; immunosuppressive; antisclerotic; cardiant
KW Virucidal; anti-AIDS; vasotropic; anti-ischaemic; antiparkinsonian;

KM anti-Alzheimer; gene therapy; human; apoptosis; fusion protein; cancer;
KM colon; breast; prostate; melanoma; lymphoma; inflammation; herpes;
KM autoimmune disorder; multiple sclerosis; viral infection.

OS Homo sapiens
yy

PN W0200056752-A2.

PD 28-SEP-2000

PF 15-MAR-2000; 2000WO-US066642.
xx

PR 24-MAR-1999; 99US-0126018.
DB 17-MAY-1999; 99US-0126030

PR 18-AUG-1999; 99US-0149449.
xxPA (HUMA-) HUMAN GENOME SCI INC
YY

PI Ruben SM, NI J, Young PA;
XY

DR WPI; 2000-587660/55.
DR N-PSDB: AA05701

XX
DT

PT the prevention, treatment and diagnosis of e.g. Alzheimer's and Parkinson's disease

XX
PS
Claim 11: Deco 354-357, 373-380, 381-384

CC The function relates to the location of genes encoding 9 human
CC apoptosis-related proteins. The nucleotide sequences AAb5570-A95798
CC can encode the human apoptosis related proteins AAb5551-B5559. The genes
CC can be used to generate fusion proteins by linking to the gene for the
CC human immunoglobulin G Fc (IgG-Fc) portion (A95759) for increasing the
CC stability of the fusion protein as compared to the human protein only.
CC The gene and encoded protein may be used in the prevention, treatment
CC and diagnosis of diseases associated with inappropriate apoptosis
CC associated protein expression, e.g. cancer (e.g. colon, breast and
CC prostate cancer, melanomas and lymphomas), inflammation, autoimmune
CC disorders (e.g. multiple sclerosis) and viral infections (e.g. herpes).

Sequence 953 AA;

Query Match	3.9%	Score 97.5;	DB 21;	Length 953;
Best Local Similarity	20.5%;	Pred. No. 3.6;		
Matches 96;	Conservative 58;	Mismatches 154;	Indels 161;	Gaps 23

```

OY 18 LKSNIIIFVFNQSTKXWALNLSISLHYNNKNSKSLFPRTPJLSKLPTELTELRKEITE 77
Db 392 lllcwlffclfcg-----llf-----raafesbpqlpdcctmclldfvllyte 430
OY 78 KIDOOIIPRPPIVHNTTTSATSHATATILNRPDYC-----RGDOLHILLEVDHL-- 127
Db 431 vhlhmppslvgrntsr-----pvelthagrdrclcagvabhmekslfvfveegqas 488
OY 128 ---GRRKQYGADPLRRMRSSPALMAGASGVTGPENNQTIVSFLFW---EGOVSLSTLL 183
Db 487 glqerdmqlg-lltra-----lpelapggddqgsyefihlltgaqfaaffllvdtdvgtqel 544
OY 182 HPSGSVALMSARNGY-----DVIYTGQFVNSTQVHSE 218
Db 542 rffqgmpppagaatscypflfpqclqgsaparedlknkhqifnll----- 592
OY 219 GGLILNTNAELCOYLDNRPDEGFYCVAPHOIMPCAL-----THYSKNKVVSYLSKO 270
Db 593 cgltskkaqkrlfhl-----vpaalarlrkrkalwblfseflr-gylsl 633
OY 271 EKSLFESNSNGVIMEKF-----NTISVSKCNTLK----- 300
Db 636 pvvqgsyefn-qvqampflimlrcelyetqsgqkvgaalrqaicanyllklylcacaaocsa 694
OY 301 -SYDLHESGLOLQHLAVLDBRN-----INIQWKYCYPLIGSMYSVKEMEYLTFRALIDTGG 355

```

Db 695 lsfvlhfpk---rlaldldnnlndyvreldpofertlvrlsvngl-----tdg 743
QY 356 GEK-----NTVYISISGQHFRRFPIDY---FIRRALNVHKAIOHLL 394
Db 744 gvkvlseeltkykivtylglyngq1-tdvgarvvtklldckqglthkl 791

Search completed: October 12, 2001, 16:03:25
Job time: 48 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 12, 2001, 16:02:37 ; Search time 16.9 seconds

(without alignments)
962,802 Million cell updates/sec

Title: US-09-729-454-1

Percent score: 2485
Sequence: 1 MKTSMINYSKILALLFLIAS.....VHPPOHYVGQINILNLYIC 475

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Sequences: 93435 seqs, 24255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1606	64.6	540	BB61_RABIT	Q05004 oryctolagus
2	107.5	4.3	4563	APP_HUMAN	P04114 homo sapien
3	104	4.2	694	VGLH_HSV60	P30001 human herpe
4	103	4.1	793	SVL_MYCPN	P73398 mycoplasma
5	101.5	4.1	1902	P2P_LACPA	Q02470 lactobacill
6	101.5	4.1	3649	ACVS_NOCIA	P27743 nocardia la
7	100	4.0	1031	TERT_EUPAE	O00939 euploies ae
8	98.5	4.0	802	CSD2_ECOLI	O26240 escherichia
9	98	3.9	1174	NOS_RHOPR	O26240 rhodnius pr
10	97.5	3.9	271	NPH1_RAT	O09366 rattus norv
11	97.5	3.9	623	ABR2_HUMAN	Q94963 homo sapien
12	97	3.9	948	P1NH_ARATH	O94963 arabidopsis
13	96.5	3.9	773	GLCB_SYNP7	P16954 synecococc
14	96	3.9	770	RPOP_AGABY3	P33539 agardicus bi
15	95.5	3.8	1102	CSOL_ECOLI	P53513 escherichia
16	95	3.8	819	POLX_TORAC	P10978 nicotiana t
17	95	3.8	1328	GRAD_MOUSE	O09311 mus musculu
18	93.5	3.8	454	DNBI_HSV60	P52338 human herpe
19	93.5	3.8	1132	YABD_SCHPO	O09853 schizosacch
20	93.5	3.8	1616	UBR1_SCHPO	O60152 schizosacch
21	93.5	3.8	1958	P1P_LACLA	P16271 lactococcus
22	92.5	3.7	1902	VGLF_MOUSE	P33481 mumps virus
23	92	3.7	538	NPH1_MOUSE	O61200 mus musculu
24	91	3.7	253	SCAA_RAT	P37089 rattus norv
25	91	3.7	698	MUTS_BUCAI	P57504 buchnera ap
26	91	3.7	1132	DNBI_HSV62	P52358 human herpe
27	91	3.7	627	CACP_CANTR	O00614 candida tto
28	90.5	3.6	1577	MISH_AGACA	P47808 acanthameb
29	90.5	3.6	1902	P2P_LACLA	P15293 lactococcus
30	90	3.6	538	VGLF_MOUSE	P19716 mumps virus
31	90	3.6	538	VGLF_MOUSE	P09458 mumps virus
32	90	3.6	538	VGLF_MOUSE	O51229 borrelia bu
33	90	3.6	538	VGLF_MOUSE	O51229 borrelia bu

34	89.5	3.6	569	Y139_MYCE	P47385 mycoplasma
35	89.5	3.6	668	PBS2_YEAST	P08018 saccharomyc
36	89.5	3.6	839	NAH3_DIDNA	O28362 didelphis m
37	89.5	3.6	1108	CN3B_RAT	O63085 rattus norv
38	89.5	3.6	1857	PAS2_PENPA	P15368 pentacillium
39	89	3.6	382	VMAT_SV41	P25182 simian virus
40	89	3.6	538	VGLF_MOUSE	P11236 mumps virus
41	89	3.6	699	SCAA_MOUSE	O61190 mus musculu
42	89	3.6	908	DPOL_BORBU	O51498 borrelia bu
43	89	3.6	958	TELL_YEAST	P38110 saccharomyc
44	89	3.6	2787	TELL_YEAST	P4641 zea mays (m
45	88.5	3.6	230	HBP_MAIZE	

ALIGNMENTS

```

RESULT 1
BB61_RABIT STANDARD: PRT: 540 AA.
AC 005004;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DR 01-FEB-1996 (Rel. 33, Last annotation update)
DE BRUSH BORDER 61.9 KDA PROTEIN PRECURSOR.
GN ADRAB-A.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
ON NCBI_TaxID=9986;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=93286138; PubMed=8509424;
RA BOLL W., Schmid-Chanda T., Senterza G., Mantel N.;
RT "Messenger RNAs expressed in intestine of adult but not baby rabbits.
RT Isolation of cognate cDNAs and characterization of a novel brush
RT border protein with esterase and phospholipase activity."
RL J. Biol. Chem. 268:12901-12911(1993).
CC -!- TISSUE SPECIFICITY: INTESTINE, AND IN A LESSER EXTENT IN KIDNEY.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE INTESTINE OF ADULT BUT NOT
CC BABY RABBIT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: Z12840; CAAT8302.1;
DR PIR: B45665; B45665.
KW SIGNAL.
FT CHAIN 23 540 POTENTIAL.
FT SEQUENCE 540 AA; 61888 MW; EB65770EC86925A CRC64;
SO
Query Match 64.6%; Score 1606; DB 1; Length 540;
Best local Similarity 57.3%; Pred. No. 6.3e-117;
Matches 309; Conservative 67; Mismatches 91; Indels 72; Gaps 2;
QY 8 YKSLALFLITASHITFTYFONSTKWSALNLSISLHYNNSTKSLPFTPLISKPLTE 67
DB 3 HNYLTLCLLAACVCIISONSSTKMGALKLNSHYNSNTSMISSIPKMSVSPVKSLE 62
QY 68 TELRIKELIEKLDQPPRPFRHYWTTSATATILNPDTYCRGDOLHILLYVRHL 127
DB 63 TELRVKELIEKLDRLPPRPFRHYWTTSATATILNPDKYCVDDOLHILLYVRHL 122
QY 128 GRRKYGDDFLRRMSSPALMAAGSGVDTFNNNGTYLVSFTLFEQGVSLILLIHPSEG 187

```

Db 123 GHOKEYGDFLRARFSPALKAGASGVTDNNNGTIVLSFTLFEQGVSLVLLHPSEG 182
 QY 188 VSALMSARNGCYDRVIFPTQGFNGTSGVHSDCLILNTAELCOYLDNRDQEGFCVPRQ 247
 Db 183 ASALMRARNGCYDRILIFKGFNGTSHVTECSLTILNSTECCYKLGDRQDFVCMKPRQ 242
 QY 248 HMFCAALTHMSKKNKVSYSLSKQESLFEPSNNGVSEMEKNTISVSKCN----- 297
 Db 243 HMFCAALTHVTSKRNKDIISLTLSKEKNLFHNSKGVSEIM-KNQHIDVSOCKSKSEYKCKQ 301
 QY 298 ----- 297
 Db 302 IGMKIPVPGGYTLQGRMLTTEPQNIQDITAKISGLCKGLIYLMGSTLRLQMTIYLPKYM 361
 QY 298 -TLKSVDLHESGKLQHLAVLDNRINICQKCPILGSMITVSKAMELIRLADRTGG 356
 Db 362 KTLKFPDLHETGFKKHLLEKHTIOKKKSHPEFTYGLFSDVHDGIIPOEDRLIG 421
 QY 357 EKNTIVIVISLGHFRFPIDVFIRBALNKAIOHLRLSPDVIKTEINIREMYNDAE 416
 Db 422 DKDTVIITFGQHFRRFPIDIFIRRAISVQAIERLFRSPATKVIKTEINIREMHIAE 481
 QY 417 RFSDFGYIOYLIDIFODLSIIDADITAGTNNVHPQHYVGNQINILLNTYC 475
 Db 482 RFGDFHGYIOYLTLDIFKDLNGVVDAMDITAIQINNHPPOVYIGNQINMFLNTYC 540
 RESULT 2
 APB_HUMAN STANDARD: PRT: 4563 AA.
 ID P04114; 000502; Q13787;
 AC 01-NOV-1986 (Rel. 03, Created)
 DT 01-NOV-1986 (Rel. 03, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE APOLIPOPROTEIN B-100 PRECURSOR (APO B-100) [CONTAINS: APOLIPOPROTEIN B-48 (APO B-48)].
 GN APOB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87016385; PubMed=3763409;
 RA Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J., Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
 RT "Complete cDNA and derived protein sequence of human apolipoprotein B-100.";
 RL Nucleic Acids Res. 14:7501-7503(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88003974; PubMed=3652907;
 RA Ludwig E.H., Blackhart B.D., Pierotti V.R., Calati L., Fortier C., Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.;
 RT "DNA sequence of the human apolipoprotein B gene.";
 RL DNA 6:363-372(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87008488; PubMed=3759943;
 RA Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H., Goto A.M., Jr., Chan L.;
 RT "The complete cDNA and amino acid sequence of human apolipoprotein B-100.";
 RL J. Biol. Chem. 261:12918-12921(1986).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87041416; PubMed=3464946;
 RA Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J., Lee N., Brewer H.B. Jr.;
 RT "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino acid sequence.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87161758; PubMed=3030729;
 RA Cladas C., Hadzopoulou-Cladas M., Nolte R.T., Atkinson D., Zannis V.I.;
 RT "The complete sequence and structural analysis of human apolipoprotein B-100: relationship between apoB-100 and apoB-48 forms.";
 RL EMBO J. 5:3495-3507(1986).
 RN [6]
 RP SEQUENCE OF 709-906 FROM N.A.
 RX MEDLINE=85270450; PubMed=3600836;
 RA Deeb S.S., Moculsky A.G., Albers J.J.;
 RT "A partial cDNA clone for human apolipoprotein B.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).
 RN [7]
 RP SEQUENCE OF 3056-3159 FROM N.A.
 RX MEDLINE=86041888; PubMed=3903660;
 RA Mehrbaban M., Schumaker V.N., Fareed G.C., West R., Chishti S.F., Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendez E., Lusis A.J.;
 RT "Human apolipoprotein B: identification of cDNA clones and characterization of mRNA.";
 RL Nucleic Acids Res. 13:6937-6953(1985).
 RN [8]
 RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.
 RX MEDLINE=86093680; PubMed=3841204;
 RA Carlsson P., Olsson S.O., Bondiers G., Darof J.S., Wiklund M., Bjursell G.;
 RT "Molecular cloning of human apolipoprotein B cDNA.";
 RL Nucleic Acids Res. 13:8813-8826(1985).
 RN [9]
 RP SEQUENCE OF 3109-4563 FROM N.A.
 RX MEDLINE=85300528; PubMed=2994225;
 RA Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F., Udea M.S., Levy-Wilson B., Powell L.M., Pease R.J., Eddy R., Nakai A., Byers M., Priestley L.M., Robertson E., Pali L.R., Bestolitz C., Snow T.B., Mahley R.W., Scott J.;
 RT "Human apolipoprotein B: structure of carboxyl-terminal domains sites of gene expression, and chromosomal localization.";
 RL Science 230:37-43(1985).
 RN [10]
 RP SEQUENCE OF 1-291 FROM N.A.
 RX MEDLINE=86149325; PubMed=3513177;
 RA Protter A.A., Hardman D.A., Schilling J.W., Miller C., Murphy T., Chen G.C., Kirshner S.W., McIntire G., Kane J.P.;
 RT "Isolation of a cDNA clone encoding the amino-terminal 291-bp of human apolipoprotein B.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).
 RN [11]
 RP SEQUENCE OF 1-1670 FROM N.A.
 RX MEDLINE=86287319; PubMed=3461454;
 RA Protter A.A., Hardman D.A., Sato K.Y., Schilling J.W., Yamanaka M., Hott Y.J., Hjertild K.A., Chen G.C., Kane J.P.;
 RT "Analysis of cDNA clones encoding the entire hApo B-100 protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).
 RN [12]
 RP PARTIAL SEQUENCE, AND IDENTIFICATION (APO-B48).
 RX MEDLINE=88018019; PubMed=3659919;
 RA Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Wong S.-H., Silbermann S.R., Cai S.-J., Deslippe J.F., Rosenbaum M., Goto A.M., Jr., Li W.-H., Chan L.;
 RT "Apolipoprotein B-48 is the product of a messenger RNA with an intron specific in-frame stop codon.";
 RL Science 238:363-366(1987).
 RN [13]
 RP DOMAINS.
 RX MEDLINE=87039351; PubMed=3773997;
 RA Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr., Innerarity T.L., Blackhart B., Taylor W.R., Marcei F., Milne R., Johnson D., Fuller M., Duets A.J., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.;
 RT "Complete protein sequence and identification of structural domains of human apolipoprotein B."

RI Nature 323:734-738(1986).
 RP [14]
 RA DOMAINS.
 RA Yang C.-Y., Chen S.-H., Gianturco S.H., Bradley W.A., Sparrow J.T.,
 RA Taitmura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M.,
 RA Lee F.-S., Gu Z.-W., Gotto A.M. Jr., Chan L.;
 RT "Sequence, structure, receptor-binding domains and internal repeats
 RT of human apolipoprotein B-100.";
 RI Nature 323:738-742(1986).
 RI [15]
 RP CALCIUM-BINDING DATA.
 RX MEDLINE-86242245; PubMed-3087360;
 RA Dashi N., Lee D.M., Mok T.;
 RT "Apolipoprotein B is a calcium binding protein.";
 RI Biochem. Biophys. Res. Commun. 137:493-499(1986).
 RI [16]
 RP VARIANT SER-4338.
 RX MEDLINE-91071750; PubMed-1979313;
 RA Navajas M., Laurent A.-M., Moreel J.-F., Ragab A., Cambou J.-P.,
 RA Cuny G., Cambien F., Rolzes G.;
 RT "Detection by denaturing gradient gel electrophoresis of a new
 RT polymorphism in the apolipoprotein B gene.";
 RI Hum. Genet. 86:91-93(1990).
 RI [17]
 RP VARIANT FDB GLN-3527.
 RX MEDLINE-89098975; PubMed-2563166;
 RA Sofia L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
 RA McCarthy B.J.;
 RT "Association between a specific apolipoprotein B mutation and
 RT familial defective apolipoprotein B-100.";
 RI Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
 RI [18]
 RP VARIANT LEU-2739.
 RX MEDLINE-91016974; PubMed-2216805;
 RA Huang L.-S., Gavish D., Breslow J.L.;
 RT "Sequence polymorphism in the human apoB gene at position 8344.";
 RI Nucleic Acids Res. 18:5922-5922(1990).
 RI [19]
 RP VARIANT FDB CYS-3558.
 RX MEDLINE-95190020; PubMed-7883971;
 RA Pullinger C.R., Hennessey L.K., Chatterton J.E., Liu W., Love J.A.,
 RA Mendel G.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
 RT "Familial ligand-defective apolipoprotein B. Identification of a new
 RT mutation that decreases LDL receptor binding affinity.";
 RI J. Clin. Invest. 95:1225-1234(1995).
 RI [20]
 RP VARIANTS L-1437; S-1914; K-2566; T-3121; A-3945; M-4128 AND T-4481.
 RX MEDLINE-97044521; PubMed-8889592;
 RA Pollier O., Ricard S., Behague I., Souriau C., Evans A.E.,
 RA Atveit D., Marques-Vidal P., Luc G., Rolzes G., Cambien F.;
 RT "Detection of new variants in the apolipoprotein B (Apo B) gene by
 RT PCR-SSCP.";
 RI Hum. Mutat. 8:282-285(1996).
 RI [21]
 RP VARIANTS FDB GLN-3527 AND CYS-3558.
 RX MEDLINE-97403938; PubMed-9259199;
 RA Rabes J.P., Vaeret M., Saint-Jore B., Erlich D., Jondeau G.,
 RA Kempf M., Giraudet P., Junien C., Bollaue C.;
 RT "Familial ligand-defective apolipoprotein B-100: simultaneous
 RT detection of the ARG3500->Gln and ARG3531->Cys mutations in a
 RT French population.";
 RI Hum. Mutat. 10:160-161(1997).
 RI [22]
 RP VARIANTS S-1914; R-1923; L-2739; D-3319; T-3427; Q-3432 AND I-3921.
 RX MEDLINE-98141125; PubMed-9490296;
 RA Leroi T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
 RT "Screening for mutations of the apolipoprotein B gene causing
 RT hypcholesterolemia.";
 RI Hum. Genet. 102:44-49(1998).
 RI [1]- FUNCTION: APOLOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF
 CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION SIGNAL
 CC FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL PARTICLES BY
 CC THE APOB/E RECEPTOR.

CC -1- DISEASE: DEFECTS IN APOB ARE A CAUSE OF FAMILIAL LIGAND-DEFECTIVE
 CC APOLOPROTEIN B-100 (FDB). IT IS A DOMINANTLY INHERITED DISORDER
 CC OF LIPOPROTEIN METABOLISM LEADING TO HYPERCHOLESTEROLEMIA AND
 CC INCREASED PRONESS TO CORONARY ARTERY DISEASE (CAD).
 CC THE PLASMA CHOLESTEROL LEVELS ARE DRAMATICALLY ELEVATED DUE TO
 CC IMPAIRED CLEARANCE OF LDL PARTICLES BY DEFECTIVE APOB/E RECEPTORS.
 CC -1- DISEASE: DEFECTS IN APOB ASSOCIATED WITH OTHER GENE DEFECTS
 CC (POLYGENIC) CAN BE THE CAUSE OF HYPOCHOLESTEROLEMIA.
 CC -1- MISCELLANEOUS: APO B-48 WHICH IS SYNTHESIZED ONLY BY THE
 CC INTESTINE, AND FOUND IN CHYLOMICRONS, IS A SHORTENED FORM OF APO
 Query Match 4.38; Score 107.5; DB 1; Length 4563;
 Best Local Similarity 20.78; Pred. No. 14;
 Matches 90; Conservative 71; Mismatches 169; Indels 105; Gaps 22;
 QY 56 KTLPLSLKPLTE-LRIKEIIEKLDQIIPPPFHV-----NTTSATRS-TATILNPRD 109
 DB 2025 KVLPLSEPINIIDLERNDAVER-----PEFTIVAFVYKNDVSNILPPFTIQ 2078
 QY 110 TYCGDGLHLLLEVDHGRKQYGD-FLRRMSPALMAAGSKVYDNNGYLVST 168
 DB 2079 EYFRNRQTIIVVENVGRNLKHNIDQFVRK-----YRAALGKLPQOAN-DYLSNFN 2130
 QY 169 LMEGQVSLLLIHPSEGVSAIWS-----ARNQGDVRF 204
 DB 2131 --MERQVS-----HAKKTLATLKRYRITENDIQALMDAKINFENKISQLOTY--MIQ 2180
 QY 205 TGQFVNGTQVHSEGGLLTNMELQYIDNDQDGEFCYCPDHPMCALTHMYSK---- 260
 DB 2181 FDQYKDSVDLHLKIALANIDILIEKKSLSDEH--YHRAVLVYTHDHLFLENID 2238
 QY 261 NKVSYLSKQKSLFERSNGVEMERFNTISVSCMTLVSDL-HESGKL-QHQLAVDL 318
 DB 2239 NKSQSSSTASMIQNDVTKYQIRIOIQKLOLK----RHIONIDIQGLAKIKHIEIDV 2294
 QY 319 DRNINIQMKCYCPPLIGSMYSVKMEHYLRAIDRTGKENYIVYSLQHPFPFP--- 375
 DB 2295 -----RYLDQLOGT-----TISFRT---NDVL-----EHVKHFVNL 2325
 QY 376 -DVFIRALNVKAIQHLRLSPDT-----WVITKTENIREYNAERPSFHYIYLI 429
 DB 2326 GDEYVAKEMKRAKVELIERVEVDQIOVLMKVELTHQYKLEKTIQKLSVLDQV 2385
 QY 430 IKDIFODLSVTIOA 444
 DB 2386 IKDYFEKLVGFIDA 2400
 RESULT 3
 VGLH_HSV6U STANDARD; PRT; 694 AA.
 AC P30001;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DE 01-OCT-2000 (Rel. 40, Last annotation update)
 DE GLYCOPROTEIN B PRECURSOR.
 GN GH OR Q48 OR LF2 OR BHLFI.
 OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6), and
 OS Human herpesvirus (type 6 / strain GS) (HHV6).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Roseolovirus.
 OX NCBI_TaxID=10370, 10369;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UGANDA-1102;
 RP STRAIN=UGANDA-1102;
 RC MEDLINE-93091236; PubMed-1333836;
 RA Gompels U.A., Carrs A.L., Sun N., Arrand J.R.;
 RT "Infectivity determinants encoded in a conserved gene block of human
 RT herpesvirus-6.";
 RI DNA Seq. 3:25-39(1992).
 RP STRAIN=UGANDA-1102;
 RC SEQUENCE FROM N.A.

[illegible]

Query Match	4.28;	Score 104;	DB 1;	Length 694;
Best Local Similarity	19.48;	Pred. NO. 1.9;		
Matches 101;	Conservative 77;	Mismatches 208;	Indels 134;	Gaps 24

0Y 17 ILASIIIFYFOKSTVMSALNLSLTHNNNSTKSFLPKPILSLKPLETELRIKETI 76
 | | | | | : | | | | | : | | | | | :
Db 2 LRLMWF-VLLPPCYGMRPLNISSHCRNGFNFPNIYRGFTTFYKNDRIRYQP 59
 | | | | | : | | | | | : | | | | | :
0Y 77 E-KLDQOIIPRPETHNTTATSTATIIL----PRDIYRCQGLIILLFVDHIGRRK 137
 | | | | | : | | | | | : | | | | | :

```

Dr 60 KCLGSDITLHLPDAINTTESLTNYEKVTRFYREPPNN----- 97
Oy 132 OYGGDGFIRAMSSPALMAGAGKATPDPNNGTLYLSTFLMEQVSLLLIHSEV--S 184
Db 98 ----DILF---SPV-----PSYKOFINDRST-----QPVVSTSLM--YSSQIYV 100
Oy 190 ALMSARNQGYDRVITFGQFVNGTSQ-----VHSEGCGLINTFNAEQLVLRKNDQYGVH 245
Db 137 RVEVROMQVDNV--SCKLPNSKELPIFYVQRK-----AKIINIVPELYTHF--I 185
Oy 246 PDM-----PCALTHMYSKNNKVASYSKQKS--LEERSVWJ-----VETEK 287
Db 186 RPFMIILYQNPADLLMYGWTGISINFRAPKPKSSFFVQTLILHLIYERKAVNYVR 245
Oy 288 FNTISYKSCITLKSVDLHESKGLQHOHLANDDRINIQ-----WKRCPLIGSR--TY 349
Db 246 FTSDDTFVDETLNDVEVALLKF-----NNLGIQLLGCQPKKINAGIYQIMPLY 298
Oy 340 SKVEMEYLTRALDRGTGKERNT-----VIVISLQHRPFPPIVFRNALNVH-----KAI 384
Db 299 GIVHESYST-----KNTGMPPLVRLVLTHTHEHLISDSVNNQCNVSEGLIYPMKK 249
Oy 390 QHLLASDPVMTIKENI-----REMNDAEFSFPGYQYONILKIDFQ 455
Db 350 ELKKEPSDSYITKNSISVSTLLYTLATAVESANTISKYMTDIAINTLQNIYKHF-- 408
Oy 436 DLVSYSIIDAMDITIAVGTNNVHPQVHVGNOINILLNYIC 475
Db 409 FTNLFSDRETLFMAETIANIIPIDERMQRHOLLIGNLC 448

RESULT 4
SYL_MYCPN
ID SYL_MYCPN STANDARD PRt: 793 AA.
AC P75398;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
Db LEUCYL-TRNA SYNTHETASE (EC 6.1.1.4) (LEUCINE--TRNA LIGASE) (LEURS).
CN LEUS OR MPN384 OR MP453.
OC Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.
OC Mycoplasmaeae; Mycoplasma.
OX NCBI_TaxID=2104;
RN RN
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129.
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Ploggens H., Pflr E., Li B.-C.,
RA Herrmann R.
RT Nucleic Acids Res. 24:4420-4449(1996).
RL Nucleonlae.
CC -I- CATALYTIC ACTIVITY: ATP + L-LEUCINE + TRNA(LEU) - AMP +
CC PHOSPHATE + L-LEUCYL-TRNA(LEU).
CC -I- SUBCELLULAR LOCATION: CYTOSOL;SMC.
CC -I- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
Dr EMBL: AE000044; AAB96101.1; -.
Dr InterPro: IPR001412; -.
Dr InterPro: IPR002300; -.
Dr InterPro: IPR002302; -.
Dr Pfam: PF00133; TRNA-synt.1; 1.
Dr PRINTS: PR00196; TRNASYNTHED.
Dr PROSITE: PS00178; AA-TRNA_LIGASE_1; 1.

```



```

RF paracasei".
CC J. Gen. Microbiol. 138:313-318(1992).
CC
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC
CC -1- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSTITUTE PREFERENCE HAVE BEEN NOTED,
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: M83946; AAA25248.1; -.
DR PIR: B44858; B44858.
DR HSSP: Q99405; IMPF.
DR MEROPS: S08.019; -.
DR InterPro: IPR000209; -.
DR InterPro: IPR001899; -.
DR Pfam: PF00746; Gram_pos_anchor. 1.
DR Pfam: PF00082; Peptidase_S8. 3.
DR PRINTS: PRO0723; Subtilisin.
DR PROSITE: PS00136; SUBTILASE_ASP. 1.
DR PROSITE: PS00137; SUBTILASE_HIS. 1.
DR PROSITE: PS00138; SUBTILASE_SER. 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
DR KMW Hydrolyase; Serine protease; Cell wall; Zymogen; Signal;
KW Transmembrane.
FT SIGNAL 1 33 POTENTIAL.
FT PROPEP 34 187 POTENTIAL.
FT CHAIN 188 1902 PIT-TYPE PROTEINASE.
FT DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1877 1895 MEMBRANE ANCHOR (BY SIMILARITY).
FT DOMAIN 1896 1902 CITOPASMIC (POTENTIAL).
FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT FT PROTEINS.
SQ SEQUENCE 1902 AA; 200253 MW; D8C9F38CE5DA582 CRC64;

Query Match 41%; Score 101.5; DB 1; Length 1902;
Best Local Similarity 21.68; Pred. No. 12;
Matches 102; Conservative 47; Mismatches 159; Indels 165; Gaps 24;

OY 58 PLISLKUTETELRIK-----ETIEKLDQIDPRPFTTHVTTSTATILNMP 108
DB 714 PAVELKMFSTDKFKFLFTNRTTIELTYQWDS-----NTDINAVYTSADPN-- 761
OY 109 DTYCRGDQHLILEVRDLGRKRGCGFLFAMSSPALMGASGKVTQFNNGTYLVSFT 168
DB 762 -----SGVLVDKKIDDAALAKGSD--ITYPAKTAQIEFT 794
OY 169 L-----FMEGVSYSLILHPSGYSALMSRNQGDRIYTGQFVNG-----TSQVHS 217
DB 795 LSLPKSTDDQQLVEGFLNKGSGD-----SLMLP--MGFGGMDNGATVDSINGITVS 847
OY 218 ECG-----LILNTNMLCOYLND--RDQEGFYCVRQHPMCAALTHYMSKNKVSYS 268
DB 848 PAGENGTVPILTKNKG--HQYYGWATDADGQIVDDQ-----ATAFSSDNALYNDIS 901
OY 269 KOEKLSPESNNGVIME-----KFTIVSVSKNTLKSVDLHESGQLQHLQDLAVDRINI 324

```

```

Db 902 M0YLLRNISNV0VLDLGGKNTLTSSSTNOTKTYDAHS----- 943
QY 325 QWQKYCY---PL-----IGSMYSKEMELLRADRTGCKNTYVI 364
Db 944 ---QKTIYNNAPAMGTYTDGRCNKTADGOSTYTLSSVP-----EGGUKRQVDEV 993
QY 365 SLGQHFRFPIDVFIRALNKHAKIOHLRLSPDTIVIIKTEINR-EMYNDAERSDFHG 423
Db 994 -----PEKLD-----SKAPYGRHVALSA-----KTENCKTQYTLTAEMKDDLSG 1032
QY 424 YIQYLIIKIDIFOD---LSVSIIDAMDITIAVGTNNVHP-----QHYVGNQIN 468
Db 1033 LDATSKVTAINEVTLNLTATFDAG--TTADGYTKIETPLSDQQAOLGNGDN 1083

RESULT 6
ACVS_NOCIA STANDARD: PRT; 3649 AA.
ID ACVS_NOCIA PRT; 3649 AA.
AC P27743;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DELTA-(L-ALPHA-AMINODIPYL)-L-CYSTEINYL-D-VALINE SYNTHETASE
DE (EC 6.3.1.1) (ACV SYNTHETASE) (ACVS).
GN PC8AB
OS Nocardi lactandurans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1913;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-VAR LC 411;
RA MEDLINE=92065808; PubMed=1956290;
RT Coque J.J.R., Martin J.F., Calzada J.G., Hiras P.;
RT "The cephamycin biosynthetic genes pcbab, encoding a large
RT multidomain peptide synthetase, and pcbc of Nocardi lactandurans are
RT clustered together in an organization different from the same genes
RT in Acetomoni chrysogenum and Penicillium chrysogenum.";
RT Mol. Microbiol. 5:1125-1133(1991).
RL
CC -1- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE
CC ACV ARE ACTIVATED AS AMINOCYL-ADENYLATES WITH PEPTIDE BONDS
CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
CC INTERMEDIATES.
CC -1- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -1- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
CC CEPHALOSPORIN.
CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X57310; CAA40561.1;
DR PIR; S18268; S18268.
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000255;
DR InterPro; IPR000873;
DR InterPro; IPR001031;
DR InterPro; IPR001242;
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00668; DUF4; 3.
DR Pfam; PF00975; Thioesterase; 1.
DR Pfam; PF00950; PP-binding; 3.
DR PRINTS; PR00154; AMPBINDING; 3.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR PROSITE; PS00435; AMP-BINDING; 1.
DR PROSITE; PS50075; ACP_DOMAIN; 3.
KW Lyase; Antibiotic biosynthesis; Multifunctional enzyme;

```

```

KW Repeat; Phosphopantetheine.
FT REPEAT 401 861
FT REPEAT 1014 1937
FT REPEAT 2079 2985
FT DOMAIN 788 857
FT DOMAIN 1864 1933
FT DOMAIN 2910 2981
FT BINDING 820 820
FT BINDING 1896 1896
FT BINDING 2944 2944
FT ACT_SITE 3502 3502
SQ SEQUENCE 3649 AA; 404079 MM; 66D095704F855E6B C0604;

Query Match
Best Local Similarity 4.18; Score 101.5; DB 1; Length 3649;
Matches 85; Conservative 54; Mismatches 141; Indels 14; Gaps 19.

QY 102 ATINPRDY--CRGDQLHILEYRDLGRKQYGGDFLARRNSPALMAG-ASQKVTIP 158
DB 1083 AGLLNKSRYRVSLLGDYD---VQHRTRKLYGLADL-----TALKAGCAMQVTV- 1180
QY 159 NNGTYLVSTFL-MEGQVSLSLIIPSEG-----VSALMSANNQGYDRIYFTQGVNGT 212
DB 1131 -----LHSLQFVWH-----KYLAIGGNTTYVGTIVSGHRLPYD----- 1166
QY 213 SQVHSEGGILINTNAELCOYLDRNDEGFCVCPQHPMCALITHYSKKKVSYLSQJER 272
DB 1167 -GIENSAGLFINT--LPILYDHQAGQNVAVNPQQAQVYTKMSK--IVHGLGQS 1220
QY 273 SFEPSNVGVEIMKENTISVSKCNLTAKSVGLDHSKGLQYLAVIDLRNINIQWQKYCP 342
DB 1221 GEKKR-----RLFDLLVLE-NYPRLLD--EEELAHQELNFEK----- 1257
QY 333 LISMTYSKEMELYLRADRTGCKNTYVIISGQHFRFPIDVF--IRNL----- 364
DB 1258 ----AYADKVDYPIAVVAREGDELTYLTMYAGELFDDIDITLLVYANTLRQVOTD 1412
QY 384 -----NVKHAIOHLRLSPDTIVIIKTEINR--- 407
DB 1113 IANPVELDLISPMKARFDSMNNEETREEPPADKTLAVFEKEMPEPEIVYRKNEL 1272
QY 408 IRMTYNDAERSDFHGYIOTL-----IINDIPQDLSYIIAMDITIAV 451
DB 1373 YRELNERANRLAHYLRVVELRPDQVALVALDKSELMITAIILAAKTKTAAY 1427

RESULT 7
TERT_EUPAE STANDARD: PRT; 1031 AA.
ID TERT_EUPAE
AC 000939;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TELOMERASE REVERSE TRANSCRIPTASE (EC 2.7.7.7) (TELOMERASE CATALYTIC
DE SUBUNIT) (TELOMERASE SUBUNIT P123).
OS Euplotis aediolatus.
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotidae;
OC Euplotes.
OX NCBI_TaxID=5940;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=97274210; PubMed=9110970;
RA Linder J., Hughes T.R., Shevchenko A., Mann M., Lundblad V.,
RA Cech T.R.;
RT "Reverse transcriptase motifs in the catalytic subunit of
RT telomerase."
RL Science 276:561-567(1997).
CC -1- FUNCTION: TELOMERASE IS A RIBONUCLEOPROTEIN ENZYME ESSENTIAL FOR
CC THE REPLICATION OF CHROMOSOME TERMINI IN MOST EUKARYOTES. IT
CC ENCATONATES TELOMERES. IT IS A REVERSE TRANSCRIPTASE THAT ADDS
CC SIMPLE SEQUENCE REPEATS TO CHROMOSOME ENDS BY COPYING A TEMPLATE
CC SEQUENCE WITHIN THE RNA COMPONENT OF THE ENZYME.

```

```

CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE REVERSE TRANSCRIPTASE FAMILY.
CC TELOMERASE SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 095964; AAC47515.1;
CC Transferase: RNA-directed DNA polymerase; Telomerase; Nuclear protein;
CC DNA-binding.
CC
CC SEQUENCE 1031 AA; 122562 MW; 57B87A63A1EED60F CRC64;
CC
Query Match
Best Local Similarity 4.0%; Score 100; DB 1; Length 1031;
Matches 106; Conservative 71; Mismatches 179; Indels 136; Gaps 29;
CC
CC 18 LASWITFTFONSTKYWSALNLSISLHANNSTKSLFPTPLISL-KPLETELRI---K 73
CC 30 LYSWJ-----QKVIKCNOSOS-HYKDEDIKIPACINIVATPRDYNEEDKVIARK 80
CC 74 EI-----IEKLDQIIPRPFTHTVNTTSATHSATILNPRDIYGRDGL---HILLEVR 124
CC 81 KVFSTGLMELIDKCL-----VELSSSDVSDROKLOCFGOLKGMQLKTHLLTA-- 131
CC
CC 125 DMLGRRKQYGGDFLRHSSPALMAGASGVYDFNNQTYLVSEFLFMGQVSLLLHP 184
CC 132 --LSTOKY--PFODEMNOVRAMIG-----NELFRHLHYKY-----LIFORT 169
CC
CC 185 SEGVSALMSAENGQYKVIETGOF---VNGTSQVH-----SEGLLIINTMAELCYDIN 235
CC 170 SECTLVQFCGNV-FDHLKYNDRKFKKKGGAADMEPRCCSTCK--YKANEDEHFLNK 226
CC
CC 246 KDQEG-----FYCVRPQHMPCALTHMYSKKKVSYLSKQKSLFERSNVGEIN 285
CC 227 INVPMNMNKKSKRTIFCYC-----THF--NRNMQFFKHE--FVSNKNNISAM 269
CC
CC 286 EKFNLT--SVSKCNTL-KSVVDHESGKLOHQLAVDDRNININOMKQYCPYLSMTYSK 342
CC 270 IRAQITITFINFRNRIRPKIKDKVIEKIAVLEKVDNFNNIYLTKSC-PLPEMREKQ 328
CC
CC 343 EMEYLTIAIDRTGGEK-----NTVIYISLGQHF-RPPIDVPIR----- 381
CC 329 KIENL---INKTREKSKYEELEFSYTDNKCVTQFINEFFYNIILPKDFLGRNRKPK 385
CC
CC 382 -----ALNVHKA1-QHLLLRSPDTMV---IKTENIREMNDARESDPHGYIYLI 430
CC 386 KYKKYVELNKKHLLKHLKINLLEKINTRFISMOVETSAKHFFYFDHENI---YVLKLL 440
CC
CC 431 KQFDULSVSI 442
CC 441 RWIFEDLVSLI 452
CC
RESULT 8
CSO2_ECOLI STANDARD: PRT: 802 AA.
AC F53513;
UT 01-OCT-1996 (Rel. 34, Created)
UT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE OUTER MEMBRANE USHER PROTEIN CSSD PRECURSOR (CS6 FIMBRIA USEH
DE PROTEIN).
GN CSSD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E8775;
RA Wolf M.K., de Haan L.A.M., Cassels F.C., Willshaw G.A.,
RA Gestel E.C.M., Gaastera W., Warren R., Boedeker E.C.;
RL Submitted (Jan-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF C6 FIMBRIAL
CC SUBUNIT ACROSS THE OUTER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC (BY SIMILARITY).
CC
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 004846; AAB51364.1;
CC InterPro: IPR000015;
CC DR Pfam: PF00577; Usher; 1.
CC DR PROSITE: PS01151; FIMBRIAL_USHER; FALSE NEG.
CC KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
CC FT SIGNAL 1
CC CHAIN 1 802 ? OUTER MEMBRANE USHER PROTEIN CSSD.
CC SEQUENCE 802 AA; 90393 MW; 5BFD0B6F17F5B5B0 CRC64;
CC
Query Match
Best Local Similarity 4.0%; Score 98.5; DB 1; Length 802;
Matches 108; Conservative 71; Mismatches 162; Indels 205; Gaps 26;
CC
CC 20 SWITFTVQ-----NSTKW-----SALULSISLHANN----- 49
CC 189 SWRLNLFEPNNDKERTWERYTYLEKSFYDKLMLVGESEYTSNYSNNYSFTGISYS 248
CC
CC 50 -----TKLEPKTPLISLPLETELRIK-----IEKLDQIIPRPFTHTVNTTSATHS 100
CC 249 TFDWYTFSELDITPELHGVADSDQIIVNGWITL--INSEVPAGFFSPITNLMYT-- 304
CC
CC 101 TATILNPRDIYCRGDQHLILEVRDHLGRKKYGGDFLRHSSPALMAGA-----S 152
CC 305 -----GGQLNV--ETDIYGNKKQY-----TVNNSLIPYMRAGLMYNFTS 344
CC
CC 153 GYVT-----DENNGTYLVSEFLFMGQVSLSL-----LIH 183
CC 345 GLTKKNSQSDDFETGQDINYGTHNS-TLFGGYQFSKNYFNLSTGIGDLSGAMLLH 403
CC
CC 184 PSE-----GVSALMSAENGQYKVIETGOFVNGTSQVHSEC 219
CC 404 VSRNSFKKRNQYINLQNTLRPNAGVNDVAYKKRYVELSDIGMHNLNOLKNSF 463
CC
CC 220 GLILNTNA-----ELCOYLDNRDQEGFYCVRP--QHPPCALTHMYSKKKVSYL 267
CC 464 SLTSLKSLNKKYNSLIDYKMKMYQWMAVDNSMSIKYFFKPR-AMITTCISLKKQSYE 522
CC
CC 268 SKQEKSLFERSNNGVEIMKEFTY-----SVSKCNT--LKSVDHESGKLOHQLAVDDR 320
CC 523 KDKKRFST--NISLPLTDYGHLSNYSFSNANTATATSSVGLNGS-----FFNDARL 573
CC
CC 321 NNIOMOK-----YCPY--LIGSMYSYEMEYLTIAIDRTGGEKNT 360
CC 574 MNIIQONRTFRNNGYTDNNTSYIATSYASPGVGTSGYSKMK---YSSQYSSASGG----- 626
CC
CC 361 VYVISLGQHRPPPIVFTFRALNHNKAIQHLILRSPDTYITENIRMT--NDAERF 418
CC 627 IVLHSDGVAFQ-----RAGDTSALVRIDNISIKIGTIPGVY 664
CC
CC 419 SDFHGY 424
CC 665 TGYNGF 670
CC

```

```

RESULT 9
NOS_RHOPR
ID AC 026240, STANDARD: PRT: 1174 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE NITRIC-OXIDE SYNTHASE, SALIVARY GLAND (EC 1.14.13.39) (NOS).
OS Rhodnius prolixus.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Palaeoptera; Hemiptera; Euhemiptera;
OC Heteroptera; Pantheroptera; Cimicomorpha; Reduviidae; Triatominae;
OC Rhodnius.
OX NCBI_TaxID=13249;
OK [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Salivary gland;
RC MEDLINE=97175053; PubMed=9022713;
RA Yuda M., Hirai M., Mura K., Matsumura H., Ando K., Chinzai Y.;
RT cDNA cloning, expression and characterization of nitric-oxide
RT synthase from the salivary glands of the blood-sucking insect
RT Rhodnius prolixus.
RL Eur. J. Biochem. 242:807-812(1996).
CC -1- FUNCTION: PRODUCES NITRIC OXIDE (NO) WHICH IS A MESSENGER MOLECULE
CC WITH DIVERSE FUNCTIONS THROUGHOUT THE BODY. THE PRODUCTION OF NO
CC IN THE SALIVARY GLAND IS USED AS A VASODILATOR FOR BLOOD SUCKING.
CC -1- CATALYTIC ACTIVITY: L-ARGININE + N NADPH + M O(2) -> CITRULLINE +
CC NITRIC OXIDE + N NADP(+).
CC -1- COFACTOR: HEME. BINDS ONE MOLE EACH OF FAD AND FMN (BY
CC SIMILARITY).
CC -1- ENZYME REGULATION: STIMULATED BY CALCIUM/CALMODULIN.
CC -1- SIMILARITY: BELONGS TO THE NOS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U59389; AAB03810.1; -
DR HSSP: P29477; 2NOS.
DR InterPro: IPR001094; -
DR InterPro: IPR001433; -
DR InterPro: IPR003097; -
DR Pfam: PF00667; FAD_binding_1.
DR Pfam: PF00175; oxidored_1ad; 1.
DR PRINTS: PR00369; FLAVODOXIN.
KW Oxidoreductase; NADP; FAD; FMN; Calmodulin-binding; Heme.
KW DOMAIN 35 45 POLY-GLN.
FT BINDING 162 162 HEME (BY SIMILARITY).
FT DOMAIN 475 495 CALMODULIN-BINDING (POTENTIAL).
FT NP_BIND 639 670 FMN (PYRIMIDINE PART) (BY SIMILARITY).
FT NP_BIND 780 791 FAD (ADP PART) (BY SIMILARITY).
FT NP_BIND 923 933 FAD (FLAVIN PART) (BY SIMILARITY).
FT NP_BIND 998 1016 NADP (RIBOSE PART) (BY SIMILARITY).
FT NP_BIND 1095 1110 NADP (ADP PART) (BY SIMILARITY).
SQ SEQUENCE 1174 AA; 132393 MW; C32F664EE51409CF CRC64;

Query Match 3.9%; Score 98; DB 1; Length 1174;
Best Local Similarity 19.4%; Pred. No. 12;
Matches 51; Conservative 53; Mismatches 77; Indels 82; Gaps 13;

```

```

DB 746 -----KVMCPILGRNJBNGSTAFATLLEIENENISV 741
QY 328 KYCYPLIGSMYTSKEN-----EYTRADITGGEKNVTVISLQHRFPPIIDVIRPA 482
DB 782 PGDH--VGVLAQNKRELVEQIISHLESAID---PDKSVOLQI-LKENTP---UGIVKRW 642
QY 383 LNVHR-----AIOHLILR-----SPTQVILIKT-----ENRMYVDAERSDFH 422
DB 833 IPHDLPLCSIRMLRFLDITPPSPPLLOFRASCATNSDQEXTLETATDSAAVEEWR 642
QY 423 GYQYLLIKQIFQDL-SVSIIDA 444
DB 893 -YWKYPNLEVEEPFVAVLPA 914

RESULT 10
NPH1_RAT
ID AC 063366, STANDARD: PRT: 271 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE NEUREXOPHILIN 1 PRECURSOR (NEUROPHILIN).
GN NPH1 OR NPH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
OK [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=96285495; PubMed=8699246;
RA Petrenko A.G., Ulrich B., Missler M., Krasnopetrov V., Rosahl T.W.,
RA Suedhof T.C.;
RT "Structure and evolution of neurexophilin.1";
RL J. Neurosci. 16:4360-4369(1996).
CC -1- FUNCTION: MAY BE SIGNALING MOLECULES THAT RESEMBLE NEUROPEPTIDES
CC AND THAT ACT BY BINDING TO ALPHA-NEUREXINS AND POSSIBLY OTHER
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE NEUREXOPHILIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L27867; AAB18420.1; -
DR Glycoprotein; Signal.
KW Glycoprotein; Signal.
FT CHAIN 1 271 POTENTIAL.
FT CARBOHYD 23 23 NEUREXOPHILIN 1.
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 271 AA; 31017 MW; 472D12D0B071E97A CRC64;

Query Match 3.9%; Score 97.5; DB 1; Length 271;
Best Local Similarity 23.7%; Pred. No. 1.7;
Matches 46; Conservative 25; Mismatches 72; Indels 51; Gaps 7;

```

3.98; Score 97.5; DB 1; Length 623;

Best Local Similarity 19.44; Pred No. 5.3; Matches 116; Conservative 82; Mismatches 172; Indels 229; Gaps 34;

```

OY 13 ALFILASWIFTFVONSTKYWSALNLSISLHYNNSTKSLFFKPL-----59
Db 72 AVGVGLAS-----HPNSTDV-HIINSLTFH-----QGLSDTKLELNSGRYGLGL 119
OY 60 -----ISLPLTELRIKEITE-KLDQOIIPRPFT-----HNITTSATISTATIL 105
Db 120 NGIGKSMLSALGKEVPIPEHIDYHILIREMPPSDKPLPHCWEVDETERALEKEERL 179
OY 106 NRPDYCRQDQHLILEYRDHLRRKQYGDPLRARBPSSPALMAGAS-----GKTDPF 158
Db 180 AHEADAC--EKLMEYERLEEL-----DADKEMRASRIILHGFTPAMQKKLKDF 229
OY 159 NNGCYLVSFTLFWECQVSL-----LLIHPSGV--SALMSARN-QGIDRY--203
Db 230 SGG-----WKRVALALALFIRPPLHLDEPTNHLDDCAVLEBELKTRILYL 280
OY 204 --FTGQFVNG--TSQVHSECGLLINTMAELCOYLNDNQ-----ESG-----241
Db 281 VSHSDPFLNGVGTNTIHH-----NKKLKYTGNGDYQVTKPLEBENGKMFHEQD 333
OY 242 -----YCVRFQH-----MPCALTLMTSKKKKSY-----266
Db 334 QIHHMKNTIARGHGSAKLARQASKEKTLQMAASGLTERVVSDDTLSTFPFGCKTTP 393
OY 267 -----LSKQESLERSNGVEIMEKFTI--SVSKCNTLK-----S 301
Db 394 PVIWQNSFKYTKDPCQIYNNEFGIDLTDRVALVGPNGAGKSTILKILGELLPTDG 453
OY 302 VDLHESGK-----LQHLAVYDLDNRNINQMAYCYPLGSGTSSVEMEYLIRADPT 354
Db 454 IRKHSVTKIGRHHQHLQEDLDLS--PLEYMAKCP-----EIKEKEMKTIIGR 503
OY 355 G--GEKNYIYA--ISLQOHR-----PPIDVIFIRALNVHNAIOLHLLRSPTMYI 402
Db 504 GLNGKQVSPIRNLSDQKCRVCLAMLANQMPHMLFLEPTN-----HLDIETDAL--555
OY 403 IKTENIREMYNDARFSDPFHGY-----QYLINDIPODLSV--SIIDAM--DITAY 451
Db 556 -----AAINEPFGKMLVSHDPRLIQVAQELVCEKQITTKPQGTI-LAY 601

RESULT 12
PINH_ARATH STANDARD: PRT: 988 AA.
AC Q9XGW1: 049256;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PINHEAD PROTEIN (ZWILLE PROTEIN).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV_LANDSBERG_ERECTA;
RC MEDLINE=96094873; PubMed=9676176;
RA Lynn K., Fernandez A., Aida M., Sedbrook J., Tasaka M., Masson P., Barton M.K.;
RT "THE PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis development and has overlapping functions with the ARGNAT1 gene.";
RL development 126:469-481(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV_LANDSBERG_ERECTA;
RC MEDLINE=96169396; PubMed=9501101;
RA Mousian B., Schoof H., Haecker A., Juergens G., Laux T.;
RT "Role of the ZWILLE gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis ";
RL

```

```

RL EMBL J. 17.1799-1809(1998).
CC -1- FUNCTION: DEVELOPMENTAL PROTEIN REQUIRED FOR RELIABLE FORMATION OF
CC PRIMARY AND AXILARY SHOOT APICAL MERISTEMS. MAY BE A COMPONENT OF
CC A HYPOTHETICAL MERISTEM FORMING COMPETENCE FACTOR.
CC -1- SIMILARITY: BELONGS TO THE ARGONAUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: AF154272; AAD40098.1; -
DR EMBL: AJ223508; CA11429.1; -
DR EMBL: Z4584; Arah:3192;24584.
DR InterPro: IPR003100; -
DR InterPro: IPR003165; -
DR Pfam: PF02170; ZAP: 1.
DR Pfam: PF02171; P4yl: 1.
KW Developmental protein.
FT CONFLICT 475 479 PROBE -> AEGOR (IN REF. 2).
FT CONFLICT 671 671 N -> D (IN REF. 2).
SQ SEQUENCE 988 AA; 110867 MW; 32EBB349C613DA20 CRC64;

Query Match 3.9%; Score 97; DB 1; Length 988;
Best Local Similarity 19.9%; Pred. No. 11;
Matches 66; Conservative 45; Mismatches 127; Indels 94; Gaps 15;

OY 212 TSOVHSEGLINTNAELCOYLDNRDQGFYCVRPQHPCALTHMYSKKNKVSLSKOE 271
DB 391 TTQPTRELMFPVDENCTMKSVIEYFOEMYGFTIOHTHLPCLQV---GNQKASYLPRMA 446
OY 272 KSLFERSNVGVEIMEKFNITISV-KCNTLRVLDHESGLOHOLAVLD----- 319
DB 447 CKIVGGRYTKRLNEKOITALKYTCQRPDRNDLRTKYOH-AKYODPYAKKEFGANIS 505
OY 320 -----RNINQOMKY-----CYPLISMTYSYKEN----- 344
DB 506 EKLASVENRILPAPWLKTHENGKEDLPQVGQNMNMKNKNGMNGMVSFACVNSRSVQ 565
OY 345 EYLTRAIDRTGKENTVIVISLQHFRRP-IDVFLRALVHKAIOHLLRSP----- 397
DB 566 ENVARGFCNELGQMEV-----SGMEFNPPEVYIPITSARPDVEALAKHYHTSMKTKGK 621
OY 398 --DMVITKTEINREMYNDAREPSDFH-GYT-QYLIIDQIFQ-----DLSYST--- 441
DB 622 EUELLALIPDNGSGTGLKRICETELGLISQCCLTTHVEKISKQYLANVSLKINVMG 681
OY 442 -----IDA-----MDI-TIAYGTNNVHP 458
DB 682 GRNTVLVDALSCRIDPLVSDIPTIFIGADYHP 713

RESULT 13
GLOB SYN7
ID GLOB SYN7 STANDARD: PRT: 773 AA.
AC P16954;
DT 01-AUG-1990 (rel. 15, Created)
DT 01-FEB-1995 (rel. 15, Last sequence update)
DE 1,4-ALPHA-GLUCAN BRANCHING ENZYME (EC 2.4.1.18) (GLYCOGEN BRANCHING
DE ENZYME).
GN GLOB.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OY NCBI_TaxId=1140;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90323609; PubMed=2142668;
RA Kiel J.A.K.M., Boels J.M., Belman G., Venema G.;

```

```

RT "Nucleotide sequence of the Synechococcus sp. PCC7942 branching
RT enzyme gene (g19b): expression in Bacillus subtilis."
RL Gene 89:77-84(1990).
CC -1- CATALYTIC ACTIVITY: FORMATION OF 1,6-GUANYLNUCLEOTIDE LINKAGES OF
CC GLYCOGEN.
CC -1- PATHWAY: THIRD STEP IN GLYCOGEN BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: THE TEMPERATURE FOR OPTIMAL ACTIVITY IS
CC APPROXIMATELY 35 DEGREES CELSIUS.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOXYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: M31544; AAB39038.1; -
DR PIR: J00550; J00550.
DR InterPro: IPR000461; -
DR Pfam: PF00128; alpha-amylase; 1.
KW Glycogen biosynthesis; Transferase; glycosyltransferase.
FT INIT_MET 0
FT ACT_SITE 439 439 BY SIMILARITY.
FT ACT_SITE 492 492 BY SIMILARITY.
FT ACT_SITE 560 560 BY SIMILARITY.
SQ SEQUENCE 773 AA; 89063 MW; 52BAAL7CA337BF57 CRC64;

Query Match 3.9%; Score 96.5; DB 1; Length 773;
Best Local Similarity 19.4%; Pred. No. 8.6;
Matches 103; Conservative 77; Mismatches 109; Indels 101; Gaps 13;

OY 33 VW-SALNLSI-SLHW-----NNSTKSLFPKTPILSKLPTETELRIKE--11 76
DB 151 VWAPSARNVSIIGDENSMDCKRHOMARNSGIMELF--IPETLYGAAYKEIKKNDDHIT 204
OY 77 EKLD-----OQIPRPPTHVNTTSAHSATILNLPRTYKGLQHLLEVRHGLNKK 141
DB 209 EKSDPFGFOEVRFP-----TASIVADLDRTYMD-----ADMLERRK 246
OY 132 Q-----YGDFLKARKSPALMAGASG-----TITNNG-----TY----- 14
DB 247 HOEELROPISVEYVHLGSMHA--SSDAIATDAQCKPLRPVPAADLKPARFLTYHELD 304
OY 164 -IYSPFLMEGGVSLSTLL--HPSEGVSAIWSANQCY-----DVIYFTGVVNTSGVH 214
DB 305 RLIPYVLDL-GYSHTELLPTAHEPFG--SMGYVGVGYAAISKRGSPETFMFVQCH 400
OY 217 SE-CGLILN-----TNAELCOYLDNRDQGFYCVRPQHPCALTHMYS 269
DB 361 QNGIGVILDMWPGHPPKDGHLAFPDGTHLYEHASRGE-----HRENGTIVENVQ 414
OY 260 KKKKVSLSKQKSLFERSNVGVEIMEKFNITISVSKNTILKSVL-----HESKRL 310
DB 413 RHEVRNPLAANLWTFDKYHI-----DGIRVDAYASMLYIDRNKKEGKILNEYGK 404
OY 311 QHOLAVLDLRNIN-----IQMORYCYPLIGSNTYSVK-FEVEYL 447
DB 465 ENLEADFLRQVNLHLSYPPGALSIAESTSWPVMSPY-----VGIGLGFWMKNMGW 520
OY 348 TRAIDRTGKENTVIVISLQHFRRP-----PIDVFLRALVHKAIOHLLRSPDTMVI 402
DB 521 HDMLD-----YFSMDPWRPQROHQNNTFSIMWAPSENMLALSH-----DEVYH 564
OY 403 IKTEINREMYNDA-EFESDFHGYIOYLLI-----KDIPODISVITDAMTI 447
DB 565 GKSNLGKMPDEWOKFANRLCLGLYMETTHGKKTLPMEGEGQMAEVNV 614

```

```

RESULT 14
ID GLGH_SYNY3 STANDARD: PRT: 770 AA.
AC P52981:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 1,4-ALPHA-GLUCAN BRANCHING ENZYME (EC 2.4.1.18) (GLYCOGEN BRANCHING
ENZYME).
GN GLGH OR SL10158.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria: Cyanobacteria: Chroococcales: Synechocystis.
OX NCBI_Taxid=1148.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugitara M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. 1. Sequence features in the 1 Mb
RT region from map positions 648 to 928 of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- CATALYTIC ACTIVITY: FORMATION OF 1,6-GLUCOSIDIC LINKAGES OF
CC GLYCOGEN.
CC -1- PATHWAY: THIRD STEP IN GLYCOGEN BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D61999; BAA10073.1;
DR InterPro: IPR000461;
DR Pfam: PF00128; alpha-amylase; 1.
KM Glycogen biosynthesis; Transferrase; Glycosyltransferase.
FT ACT_SITE 433 433 BY SIMILARITY.
FT ACT_SITE 486 486 BY SIMILARITY.
FT ACT_SITE 554 554 BY SIMILARITY.
SQ SEQUENCE 770 AA: 89527 MW: A435ACAT7703FABA CRC64;

Query Match 3.94; Score 96; DB 1; Length 770;
Best local similarity 18.74; Pred. No. 9.3;
Matches 102; Conservative 84; Mismatches 201; Indels 158; Gaps 29;

07 33 VM--SALNLSI--SLHYW-----NSTRKSLFRTPLISLKLTELRKE--II 76
08 146 VMAPNARNYSLIGDFNMKDGRLHOMKRNMMWELF--PELVGTSYVEIKKMEGHII 203
09 77 EKLD-----QQLPRPTHTVNTTSATNSTATILNPRDYCRDQHLILLEV--DHLG 128
10 204 EKTDPYGFYQEVRRK-----TASIVADLDGVOMHDE--DWLEARRTSPLS 247
11 129 R---RKQYGDPLFARMSSPALMAGAS--KYLDNNGTIVYSTLFW----- 172
12 248 KPVSYVELHLSWMLHTAYDEPVYTLTGEGVEYSENNMG--ARFLTYELVYDKLPIYV 304
13 173 ----GVSLSLIT--HPSEVSALMSARNQY-----DRYIFNGQ-FVNGTSQV 215
14 405 KRLATYTHIELFLIAEHPFDG---SMGYQVGYAPYTRFSGPEDFMYFDQCHLNGIGVI 361
15 216 -----HSEGLILNTNAELCOYLNRDQSEFYVRRQHPALATHMYSNKK 263
16 362 IDWVGFHEPKDGH---GLAFDQTHLYHEGPR-----KQEHKMGTLIFNYGRNEY 410
17 264 VSYLSKQKSLFERSNV--GVEIMKFNITSVSKCNLTLSVDLHESGKLOHOLAVDLDRNT 322

```

```

DB 411 RNELVANALWFEDKTHIDMRVDAVASMLYDCREGEPMVANEYGRNLEAFADFLRQV 470
09 323 N-----IOMKQCYCPILGSMYTSYK-EMELTLAIDRTGGEKN 359
10 471 NSVSYTPPGILSLAEESTSMWNSMPT-----VGLGKNNLKNMGMHMDLD----- 519
09 360 TYIVISLQGFRRP-----PIDVFLRALNVAKIQHLLSPDPTWYIITENIREMYND 414
10 520 ---VPSMDPFRQFHONSITFSMWNSSENYMLALSH-----DEVHGKSNMLKMPGD 570
09 415 A-EFSPDFHYIYLLI-----KDFQDLSIIDDADITLIVAGTNNHPPQHYVGNQINI 469
10 571 EMQKANYRALFTYMTHTGKKTMPMSMERGQSENNWAGDLEWHLLNFPHQOLKQFFT 630
09 470 LNTYI 474
10 631 ELNHL 635

RESULT 15
RPOF AGABT
ID RPOF AGABT STANDARD: PRT: 1102 AA.
AC P33539:
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PROBABLE DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6).
OS Agaricus bisporus.
OG Mitochondrion.
OG Plasmid pEM.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Agaricales;
OX Agaricaceae; Agaricus.
OX NCBI_Taxid=5343;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 24666 / AG4;
RX MEDLINE=91347410; PubMed=1879001;
RA Rodison M.M., Boyer J.C., Horgen P.A.;
RT "Homology between mitochondrial DNA of Agaricus bisporus and an
RT internal portion of a linear mitochondrial plasmid of Agaricus
RT bisporus.";
RL Curr. Genet. 19:495-502(1991).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
CC RNA(N)
CC -1- SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63075; CAA44799.1;
DR PIR: S28104; S28104.
DR InterPro: IPR002092;
DR Pfam: PF00940; RNA_POL_1.
DR PROSITE: PS00489; RNA_POL_2_1.
DR PROSITE: PS00900; RNA_POL_PHASE_1_1.
KM Transferase; Transcription; DNA-directed RNA polymerase;
KM Mitochondrion; Plasmid.
FT ACT_SITE 734 734 BY SIMILARITY.
FT ACT_SITE 804 804 BY SIMILARITY.
FT ACT_SITE 980 980 BY SIMILARITY.
SQ SEQUENCE 1102 AA: 126664 MW: F72FDD308D1AB5B4 CRC64;

Query Match 3.88; Score 95.5; DB 1; Length 1102;

```

